Previsão Probabilística dos Desvios dos Agentes Comerciais e Produtores do Mercado de Eletricidade

Tese submetida à Faculdade de Ciências da
Universidade do Porto para Obtenção do grau de Mestre
em Engenharia Matemática

Aluno

Cézar Vicente Cerciari

Orientador

Doutor Ricardo Jorge Gomes Sousa Bento Bessa

Departamento de Matemática

04 de julho de 2017

Agradecimentos

Antes de mais nada, quero agradecer não apenas a uma pessoa ou um grupo de pessoas, mas sim, quero agradecer a Portugal. Sei que não é comum agradecer um país, mas não posso deixar de me sentir extremamente grato a esse país que me recebeu de braços abertos, que me ensinou tanto e que me fez sentir em casa.

Também quero agradecer a essa nova família e amigos que fiz aqui.

Os meus sinceros agradecimentos ao meu orientador Doutor Ricardo Bessa, por toda a disponibilidade, conselhos e críticas construtivas. E a todos meus professores desse Mestrado, que me ensinaram tanto.

Tenho muito a agradecer a minha família e amigos no Brasil, que mesmo longe, sempre me apoiaram e motivaram nessa etapa tão importante da minha vida.

E claro, quero agradecer muito a minha namorada Sara que sempre esteve do meu lado, me apoiando e nunca deixando que eu me desmotivasse nem mesmo perante as maiores dificuldades.

Resumo

O MIBEL – Mercado Ibérico de Eletricidade, é um mercado regional de energia elétrica entre Portugal e Espanha. Com esse mercado é possível que os consumidores instalados no espaço ibérico comprem energia elétrica com livre concorrência a qualquer agente produtor ou comercializador de energia elétrica.

Para controle e segurança do sistema elétrico, cada Agente realiza previsões de geração, compra e venda de energia para as próximas horas/dias.

O operador do sistema elétrico REN (Redes Energéticas Nacionais) é responsável pela gestão do Sistema Elétrico e necessita prever a distribuição de probabilidade condicionada dos desvios dos Agentes.

Nesta dissertação é analisado e desenvolvido diferentes modelos de previsão probabilística para o desvio Total e por Unidade de Agente: Regressão Linear de Quantis, *Quantile Regression Forest* e *Gradient Boosting Machine*.

Para cada um dos três modelos apresentados, serão analisados os Erros médios associados assim como as avaliações de desempenho para cada método de previsão.

Palavras-chave: Previsão Desvio de Energia Elétrica, Previsão Probabilística, Regressão de Quantis, Desvio Agentes Comercializadores e Produtores.

Abstract

The MIBEL - Iberian Electricity Market, is a regional electricity market between Portugal and Spain. With this market is possible that the installed consumers in the Iberian space buy electric energy with free competition to any producer or trading of electric energy.

For control and security of the electrical system, each Agent makes forecasts of generation, purchase and sale of energy for the next hours/days.

The electrical system operator REN (National Energy Networks) is responsible for the management of the Electrical System and needs to predict the conditional probability of the agent's deviations.

In this dissertation is analyzed and developed several probabilistic prediction models for the Total and Agent Unit deviation: Quantile Linear Regression, Quantile Regression Forest and Gradient Boosting Machine.

For each of the three presented models, will be analyze the associated mean errors as well as the evaluations performace for each forecasting method.

Keywords: Electric Power Deviation Forecasting, Probabilistic Forecasting, Quantis Regression, Deviation Trading and Producers Agents.

Sumário

Lista de Tabelas

Lista de Figuras

Abreviaturas e Siglas

REN – Redes Energéticas Nacionais

MIBEL – Mercado Ibérico de Eletricidade

GBM – Gradient Boosting Trees

PHF – Programa Hora Final

MAE - Mean Absolute Error

RMSE - Root Mean Square Error

MAPE - Mean Absolute Percentage Error

1 Introdução

Os agentes produtores e comercializadores de energia elétrica apresentam ofertas de compra e venda (quantidade e preço) para as próximas horas/dia do mercado Ibérico de eletricidade. Estas ofertas são definidas com base em previsões de consumo e produção de energia elétrica e num racional económico. Neste contexto, o operador do sistema elétrico (REN) necessita de prever a distribuição de probabilidade condicionada dos desvios dos agentes com o objetivo de estimar o risco operacional do sistema elétrico.

No ano 2000 a REN, devido a um processo de privatização, foi separada do Grupo EDP [26] e era responsável pelo transporte de energia elétrica e gestão do sistema elétrico. Em 2007, com a nova reestruturação onde incluía a aglomeração das infra-estruturas relativas ao gás natural, foi alterada de Rede Eléctrica Nacional para Redes Energéticas Nacionais.

O MIBEL (Mercado Ibérico de Electricidade), é um acordo entre Portugal e Espanha para a constituição de um mercado regional de electricidade, contemplando a liberdade de acesso de todos os agentes a todas as plataformas de negociação, promovendo assim um regime de livre concorrência.

Para cada agente de energia elétrica pode haver uma ou mais unidades, e cada uma dessas unidades apresenta a previsão de compra e venda de energia elétrica para as próximas horas e dias, que será comparado com o valor real futuramente. Por tratar-se de uma previsão, ocorrem desvios entre o valor previsto pelas unidades e o valor real. Esses desvios serão analisados e realizado métodos para prevê-los.

1.1 Motivação

É de fundamental importância que o Sistema de Energia Elétrica seja controlado e seguro. Métodos de previsão probabilística devem ser estudados e utilizados para assegurar que o Sistema não seja sobrecarregado.

Ao longo desta Dissertação, serão referenciados muitos trabalhos onde foram estudados o mercado de eletricidade em Portugal, porém, em sua grande maioria são estudos que tem como objetivo realizar previsões para o preço da energia elétrica e não a quantidade de venda e compra. Ou seja, a necessidade de estudar os desvios dos agentes tornam-se ainda mais necessários.

1.2 Visão Geral

Na década de 90 deu inicio a reestruturação dos mercados de eletricidade na maioria dos países da Europa Continental, sendo que em 2007 foi dado início ao Mercado Ibérico de Eletricidade (MIBEL) [24].

Ao longo dos anos foram surgindo novos Agentes Comercializadores e Produtores de energia elétrica, sendo que em 2017 o total são em 35 Agentes Comercializadores e 3 Produtores [12] conforme tabelas 1 e 2.

Cada Agente tem a liberdade de comprar e vender energia elétrica num regime de leilão [17], o que ocasiona em um mercado onde torna-se necessário a previsão não apenas de preços mas também de quantidade.

	Código	Designação
1	AUDAX	Audax Energia, S.L.
2	AUDPT	Audax Energia, S.L. - Sucursal Portugal
3	ECOCH	ECOCHOICE, S.A.
4	EDFT	EDF Trading Limited
5	EDP	EDP Comercial - Comercialização de Energia, S.A.
6	EDPSU	EDP Serviço Universal
7	EGED	ACCIONA Green Energy Developments, S.A.
8	EGLE	AXPO IBERIA, S.L
9	ELERG	Elergone Energia, Lda.
10	ELUSA	ELUSA, Lda.
11	ENAT	ENAT - Energias Naturais, Lda
12	ENDCO	Endesa Energia, S.A.
13	ENDG	Endesa Generación, S.A. (Comercializador)
14	ENDP	Endesa Energia, S.A. - Sucursal de Portugal
15	ENFOR	ENFORCESCO, S.A.
16	EYGAS	ElyGas Power, S.L.
17	FORTI	Fortia Energia, S.L.
18	GALPW	Galp Power, S.A.
19	GASN	Gas Natural SDG, S.A.
20	GNCO	Gas Natural Comercializadora S.A.
21	GNSE	Gas Natural Servicios SDG, S.A.
22	GOLDE	Goldenergy - Comercializadora de Energia, S.A.
23	HENSE	HEN - Serviços Energéticos, Lda.
24	IBCLI	Iberdrola Clientes, S.A.
25	IBCOM	Iberdrola, S.A.
26	IBGES	Iberdrola Generación España, S.A.
27	IGES	Iberdrola Generación - Energia e Serviços Portugal
28	JAFP	JAFPlus, Lda.
29	LOGIC	Logica Energy, Lda.
30	LUZBO	LUZBOA - Comercialização de Energia, Lda.
31	NEXU	Nexus Energía, S.A.
32	PHENE	PH Energia Unipessoal, Lda.
33	ROLEA	Rolear - Automatizações, Estudos e Representações, S.A.
34	VIECO	E.ON Generación, S.L.

	Código	Designação
1	EDPGP	EDP - Energias de Portugal, S.A.
2	EGEN	Endesa Generación, S.A.
3	RENTR	REN - Trading, S.A.

Tabela 2: Agentes Produtores

1.3 Objetivos dos Estudos

O objetivo deste trabalho consiste na aplicação de modelos estatísticos para previsão probabilística das séries temporais dos desvios totais e por agente.

Os desvios são a diferença entre a quantidade prevista e a quantidade real de compra ou venda de energia elétrica, e dependem de diversos fatores como o dia, hora, geração de energia, previsão de carga total, entre outros.

Tendo em vista os fatores que influenciam os desvios, serão utilizados métodos de previsão probabilística para estimar os desvios tanto de uma forma geral como de cada unidade individualmente.

Por fim, serão comparados os modelos de previsão para então ser determinado qual o melhor método para esse cenário.

1.4 Problema

O MIBEL oferece um mercado livre de comercialização de energia elétrica, o que resulta na existência de diversos Agentes Produtores e Comercializadores. Cada Agente realiza previsões de produção, compra e venda de energia elétrica por hora/dia, porém há desvios que devem ser minimizados.

Todo o sistema deve ser monitorado afim de prevenir qualquer problema que possa comprometer a rede elétrica.

Para isso é de extrema importância a aplicação de métodos de previsão para assegurar a confiabilidade do sistema elétrico.

1.5 Estrutura da Dissertação

A presente dissertação está dividia em quatro partes. A primeira parte é uma introdução ao que será estudado. Na segunda parte é apresentada a fundamentação teórica juntamente com as referências. Na terceira parte é mostrada a metodologia utilizada desde a criação da base de dados

até os modelos estatísticos. Por fim, a quarta parte apresenta a implementação prática dos modelos de previsão, os resultados e a avaliação de desempenho de cada modelo.

2 Estado da Arte e Fundamentação Teórica

Na Faculdade de Engenharia da Universidade do Porto já foram realizadas dissertações com estudos de previsão probabilística para o mercado de eletricidade.

Os trabalhos [2] [17] utilizam técnicas de regressão de quantis e *Gradient Boosting* para prever o preço da energia elétrica para cada hora do dia no MIBEL.

Também visando a previsão probabilística dos preços de energia elétrica no MIBEL, o trabalho [9] utiliza a ferramenta de previsão NW-KDE.

Em [24] são fornecidas muitas informações da reestruturação do Setor Elétrico na Europa, além de uma análise sobre Redes Neuronais Artificiais para previsão de preços de energia elétrica assim como em [25], [26] e em [27]

A dissertação [28] utiliza método de previsão tecnologia para a previsão dos preços da energia a longo prazo.

Porém, todos os estudos citados acima tem como objetivo a previsão do preço da energia elétrica no MIBEL, e o foco da presente dissertação é a previsão da quantidade de energia elétrica, mais precisamente dos desvios da compra e venda.

As fontes para a criação da base de dados foram [11] [12].

2.1 Modelos Estatísticos

Os modelos estatísticos que serão analisados e utilizados são métodos baseados em Machine Learning, onde consiste na previsão probabilística de uma variável (Target) que depende de um vetor de variáveis explicativas. O Machine Learning é capaz de adquirir conhecimento de forma automática que toma decisões baseado em experiências acumuladas [6].

2.1.1 Regressão Linear Múltipla

Os métodos de regressão são a escolha padrão para analisar e descrever a relação entre uma variável resposta e um conjunto de variáveis explicativas [29], e apesar do grande número de alternativas que já existem para previsão, os métodos de regressão linear continuam a ser muito utilizados [17].

O modelo de Regressão Linear Múltipla descreve uma relação entre um conjunto de variáveis explicativas X_i e uma variável dependente Y [19] [20].

$$Y_i = \beta_0 + \beta_1.x_{1i} + \beta_2.x_{2i} + ... + \beta_k.x_{ki} + \varepsilon_i, i = 1, 2..N \tag{2.1.1}$$

Onde: β_i são o conjunto de parâmetros que relacionam a resposta Y com as variáveis de regressão X_i.

ε_i é o erro aleatório associado ao valor observado y_i.

Para estimar os parâmetros β_i, é utilizado o método dos mínimos quadrados.

$$\hat{\beta} = ([x]^T.[x])^{-1}.[x]^T.y \tag{2.1.2}$$

Para calcular o y estimado $\hat{y}$ é utilizada a Equação 2.1.3.

$$\hat{y} = [x].\hat{\beta} \tag{2.1.3}$$

O erro associado é a diferença entre o valor valor real e o valor estimado:

$$\varepsilon = y - \hat{y} \tag{2.1.4}$$

Para avaliar a adequação do modelo e das variáveis de regressão à resposta é realizado um conjunto de testes.

Para avaliar o significado da regressão e verificar se existe uma relação forte entre a resposta e as variáveis explicativas é calculado o coeficiente de determinação R^2.

$$R^2 = \frac{SS_{reg}}{SS_y} = 1 - \frac{SS_{erros}}{SS_y} \tag{2.1.5}$$

Onde:

$$SSy = (y - \bar{y})^T(y - \bar{y}) = SS_{erros} + SS_{reg} \tag{2.1.6}$$

$$SSreg = (\hat{y} - \bar{y})^T(\hat{y} - \bar{y}) \tag{2.1.7}$$

O coeficiente de determinação está compreendido em um valor entre 0 e 1. Quanto mais próximo de 1 mais as variáveis explicam a resposta, consequentemente, melhor é o modelo.

Outro teste que deve ser realizado é a analise da significância da regressão, que avalia se existe uma relação linear entre as amostras e as variáveis de regressão. É então realizado um teste de

hipótese, onde a hipótese nula diz que todos os coeficientes de regressão $(\beta_1, \beta_2, ..., \beta_k)$ são nulos; e a segunda hipótese é que existe pelo menos um coeficiente de regressão não nulo.

$$F_0 = \frac{\frac{SS_{reg}}{k}}{\frac{SS_{erros}}{n-(k+1)}} \tag{2.1.8}$$

Quando F_0 é superior a $f_{\alpha,k,n-(k+1)}$ [1] rejeitamos a hipótese nula e o modelo é considerado adequado.

Também é necessário verificar a adequação das variáveis de regressão, realizando um teste a cada coeficiente de regressão comparando-o com a função de distribuição T de Student.

Para ser rejeitada a hipótese nula é necessário que $\mid T_0 \mid > \frac{t_\alpha}{2}, n-(k+1)$

Onde:

$$T_0 = \frac{\hat{\beta}_j}{\sqrt{\sigma^2.C_{jj}}} \tag{2.1.9}$$

$$C = (X^T.X)^{-1} \tag{2.1.10}$$

$$\sigma^2 = \frac{SS_{erros}}{n-(k+1)} \tag{2.1.11}$$

Em que $n-(k+1)$ representam o número de graus de liberdade e σ^2 é a variância.

Pode ser vantajoso tratar a incerteza do modelo de forma separada e sem assumir algum tipo de distribuição logo à partida, e portanto não influenciar nem os resultados das previsões nem o erro destas durante o processo de previsão. Para tal, pode-se fazer uma regressão por quantis.

2.1.1.1 Regressão Linear de Quantis

Os quantis são uma medida estatística que quantifica um conjunto de dados [17]. Os quantis de uma amostra são por vezes expressos na forma $\tau \in [0,1]$, então para o τ-ésimo quantil, 100x τ % das observações deverão ter um valor inferior ao quantil τ.

Em 1978, Roger Koenker e Gilbert Bassett apresentaram um modelo estatístico para a estimação de quantis chamado Regressão Linear de Quantis [5].

[1] Valor obtido a partir dos valores tabelados da função estatística F, ou função Fisher.

$$F(y \mid X = x) = P(Y \leq y \mid X = x) \tag{2.1.12}$$

Onde Y é uma variável aleatória e x é um vetor de variáveis explicativas.

O quantil condicional [2] é definido por:

$$Q(\tau, x) = inf\{y : F(y \mid X = x) \geq \tau\} \tag{2.1.13}$$

Onde $0 \leq \tau \leq 1$

O quantil condicional de ordem τ[2] pode ser expresso como uma combinação linear das variáveis explicativas:

$$Q(\tau, X) = \beta_0(\tau) + \beta_1(\tau)x_1 + ... + \beta_p(\tau)x_p \tag{2.1.14}$$

Em que x são as variáveis explicativas e $\beta(\tau)$ são coeficientes desconhecidos que dependem de τ.

2.1.2 Random Forest

Random Forests é um método de previsão composto por uma coleção de árvores de decisão que serão usadas para classificar um novo exemplo por meio do voto majoritário[6]. São criadas amostras aleatórias de conjuntos de treinamento onde cada novo conjunto é construído a partir do conjunto de treinamento original. Um subconjunto de *m* atributos é selecionado aleatoriamente e avaliado a cada nó da árvore.

2.1.2.1 Quantile Regression Forest

Tendo como base o modelo Random Forest, o Quantile Regression Forest[7] [16] é um método de regressão por quantis. A média condicional $E(Y \mid X = x)$ é estimada pela média das previsões das árvores aleatórias.

$$\omega_i(X) = k^{-1}\sum_{t=1}^{k}\omega_i(X, \theta_t) \tag{2.1.15}$$

Onde $\omega_i(X)$ é a média dos $\omega_i(\theta)$ das k árvores, e a previsão da random forest é dada por:

$$\hat{\mu}(X) = \sum_{i=1}^{N}\omega_i(X)y_i \tag{2.1.16}$$

2.1.3 GBM - Gradient Boosting Machines

O algoritmo de aprendizagem automática Gradient Boosting Machines pode ser aplicado em problemas de regressão ou classificação[4]. Gradient Boosting foi criado por Jerome H. Friedman em 1999 [3] [18]. O método *gradient boosting* consiste na minimização de uma função de custo que penaliza a diferença ente os valores obtidos pelo modelo preditivo e os valores medidos [17]. Esse método de previsão consiste em uma variável aleatória de resposta y (Target) e um vetor de variáveis explicativas $x = \{x_1, ..., x_n\}$ Utilizando uma base de dados de treino $\{y_i, x_i\}_1^N$ com os valores de (y, x) conhecidos. O objetivo é obter uma aproximação de $\hat{F}(x)$, da função $F*(x)$ mapeando x para y que minimiza o valor esperado da função de perda $L(y, F(x))$ através da função de distribuição de todos os valores de (y, x) [3]

$$F* = arg\min_{F} E_{y,x}L(y, F(x)) = arg\min_{F} E_x[E_y(L(y, F(x))) \mid x] \tag{2.1.17}$$

2.2 Avaliação do Desempenho

Para analisar o Erro gerado nos modelos de previsão, é utilizado Mean Absolute Error (MAE)[2], Root Mean Square Error (RMSE) e Mean Absolute Percentage Error (MAPE).

$$MAE = \frac{1}{N}\sum_{t=1}^{N} \mid y_t - \tilde{y}_t \mid \tag{2.2.1}$$

Onde N é a quantidade observada, y_t é o valor real no instante t e $\tilde{y}_t$ é o valor previsto no instante t.

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \tilde{y}_t)^2} \tag{2.2.2}$$

Onde N é a quantidade observada, y_t é o valor real no instante t e $\tilde{y}_t$ é o valor previsto no instante t.

$$MAPE = \frac{100}{T} \sum \frac{Y(t) - \hat{Y}(t)}{Y(t)} \tag{2.2.3}$$

Onde $Y(t)$ – corresponde à sucessão cronológica univariada;

$\hat{Y}(t)$ – corresponde ao valor estimado pela análise de sucessões cronológicas;

T – corresponde ao número total de observações utilizadas.

2.2.1 Calibration

Calibration (ou *Reliability*) [14] [23] avalia o quanto os quantis estimados diferem dos quantis nominais [13]. Para essa estimação, é necessário calcular o *Calibration* para cada um dos quantis estimados. Para isso é necessário introduzir uma variável indicadora $\xi_{t,k}^{\alpha}$, que é calculada a partir de um quantil estimado $\hat{q}_{t+k|t}^{(\alpha)}$ e o valor real p_{t+k} no instante t para $t+k$,

$$\xi_{t,k}^{\alpha} = 1_{t+k<\hat{q}_{t+k|t}^{(\alpha)}} = \begin{cases} 1, & \text{se } p_{t+k} < \hat{q}_{t+k|t}^{(\alpha)} \\ 0, & \text{se ao contrário} \end{cases} \tag{2.2.4}$$

A variável indicadora $\xi_{t,k}^{\alpha}(t = 1, ..., N)$ é uma sequencia binária que indica se o valor real p_{t+k} está abaixo do quantil estimado.

O estimador para a cobertura atual $a_k^{\alpha} = E[\xi_{t,k}^{\alpha}]$ que é obtido pelo calculo da média de $\xi_{t,k}^{\alpha}$:

$$\hat{a}_k^{\alpha} = \frac{1}{N} \sum_{t=1}^{N} \xi_{t,k}^{\alpha} = \frac{n_{k,1}^{\alpha}}{n_{k,0}^{\alpha} + n_{k,1}^{\alpha}} \tag{2.2.5}$$

É possível obter o diagrama de *Calibration* onde é comparado o *Calibration* observado e o *Calibration* perfeito. O desvio a um *Calibration* ideal b_k^{α} é dado por:

$$b_k^{\alpha} = \alpha - \hat{a}_k^{\alpha} \tag{2.2.6}$$

O diagrama de *Calibration* tem como objetivo apresentar o desvio do Calibration observado e o *Calibration* ideal para cada quantil nominal.

Sendo 0 o *Calibration* ideal, quando mais próximo da diagonal o diagrama for, melhor.

A Figura 1 mostra um exemplo de diagrama *Calibration* para dos quantis entre 5% e 95% [14].

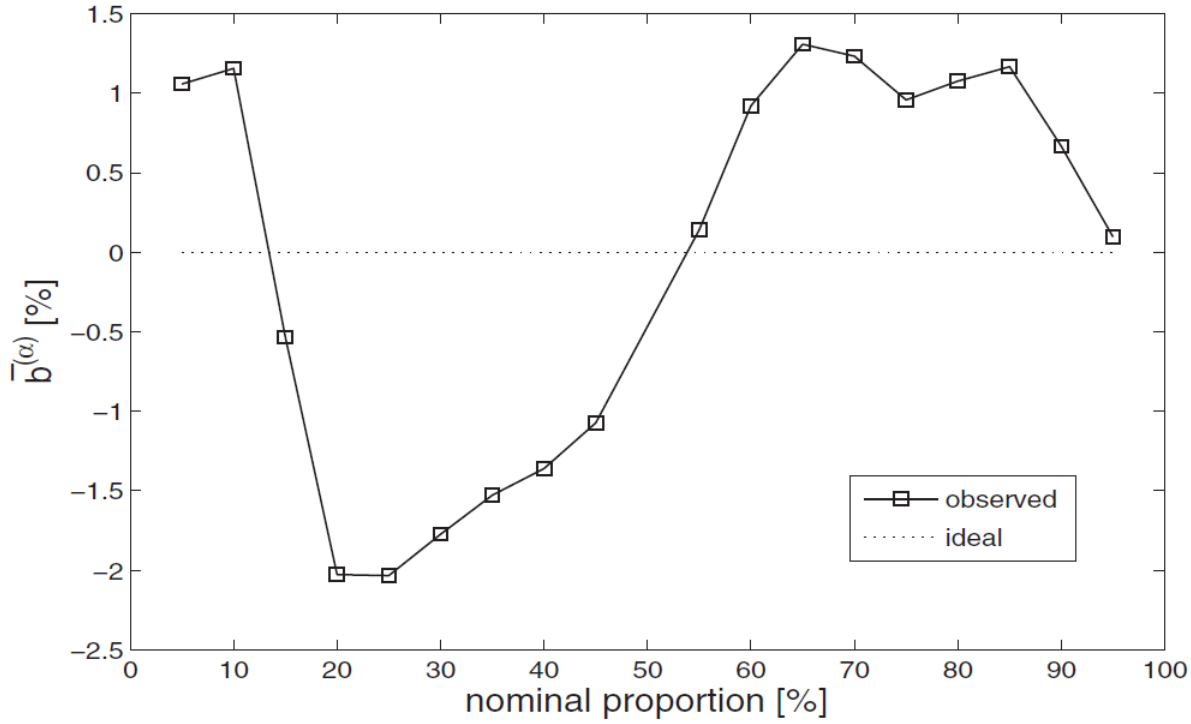

Figura 1: Exemplo de Diagrama Calibration

Onde $\bar{b}^{(\alpha)} = 1/k_{max} \sum_k b_k^{(\alpha)}$

2.2.2 Sharpness

Sharpness representa a capacidade do modelo de previsão em prever acontecimentos de uma forma precisa [9], e ao contrario do *Calibration*, é calculado independentemente da comparação dos valores observados na previsão e os valores reais, avaliando a distância entre dois quantis, fornecendo informação sobre a forma da distribuição [2].

Sharpness é calculada para pares de quantis [14]:

$$\delta_{t,k}^{\beta} = \hat{q}_{t+k|t}^{1-\frac{\beta}{2}} - \hat{q}_{t+k|t}^{\frac{\beta}{2}} \tag{2.2.7}$$

Com taxa de cobertura nominal $1 - \beta$, a equação 2.2.7 calcula o tamanho do intervalo estimado no instante t para $t + k$.

O Sharpness para esses intervalos e para o horizonte k é dado por $\bar{\delta}_k^{\beta}$ que é o tamanho médio dos intervalos.

$$\bar{\delta}_k^{(\beta)} = \frac{1}{N} \sum_{t=1}^{N} \delta_{t,k}^{(\beta)} = \frac{1}{N} \sum_{t=1}^{N} (\hat{q}_{t+k|t}^{1-\frac{\beta}{2}} - \hat{q}_{t+k|t}^{\frac{\beta}{2}}) \tag{2.2.8}$$

Portanto, Sharpness não pode ser calculada para um único quantil, mas sempre para pares de quantis.

Para visualizar a avaliação do Sharpness, é possível utilizar o δ-Diagrams que é criado a partir da função da taxa de cobertura nominal dos intervalos $\bar{\delta}_k^{(\beta)}$
$\bar{\bar{\delta}}^{\beta} = 1/k_{max} \sum_k \bar{\delta}_k^{(\beta)}$

2.2.3 Skill Score

A qualidade da previsão probabilística não pode ser avaliada apenas pelo *Calibration* e *Sharpness* separadamente, mas sim pelas duas em conjunto [9].

As *skill score* consistem e um único critério de avaliação que contém informação relativa as propriedades *Calibration* e *Sharpness*.

Essa avaliação é dada pelas regras de pontuação que associal um único valor numérico $Sc(\hat{f}, p)$ [14] para uma distribuição $\hat{f}$ se o evento p se materializar.

$$Sc(\hat{f}', \hat{f}) = \int Sc(\hat{f}'(p), p)\hat{f}(p)dp \tag{2.2.9}$$

Mesmo sendo que *Calibration* e *Sharpness* são propriedades intuitivas e que são facilmente interpretadas com diagramas, eles apenas contribuem para uma avaliação diagnóstica do método. Eles não permitem concluir a qualidade do método como é o caso das *skill score* que pode ser calculada pela equação 2.2.10.

$$Sc(\hat{f}, p) = \sum_{i=1}^{m} (\xi^{(\alpha_i)} - \alpha_i)(p - \hat{q}^{(\alpha_i)}) \tag{2.2.10}$$

Este resultado é positivamente orientado e admite um valor máximo de 0 para previsões probabilísticas perfeitas

3 Metodologia

Tanto para a criação da base de dados, testes, desenvolvimento, implementação prática e análise dos resultados, foi utilizado o software estatístico R [22].

Antes de ser formada a metodologia à ser seguida para a presente dissertação, foi analisado o cenário geral da venda/consumo de energia elétrica em Portugal.

3.1 Análise dos Dados

Utilizando o banco de dados da venda de energia por Agente[12] foi possível analisar a venda total de energia de 2015, assim como analisar por mês e por Agente.

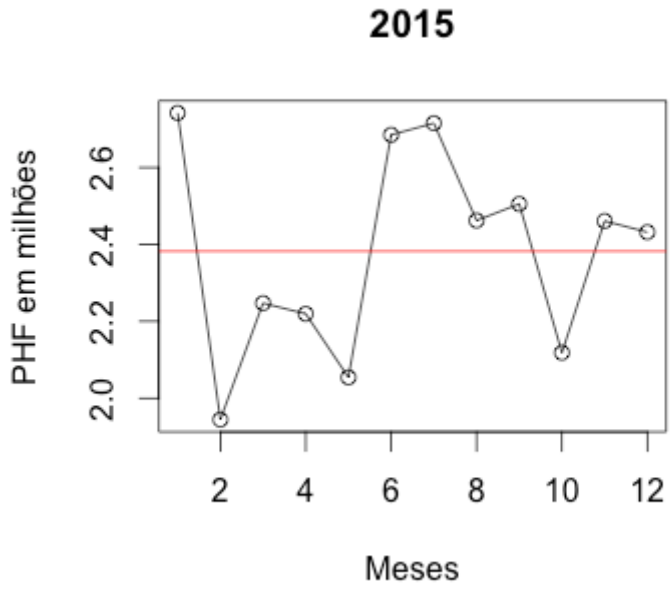

Figura 2: Consumo/Venda total de energia em 2015

Como é possível observar na Figura 2, os meses de janeiro, junho e julho são os que mais se consome energia elétrica, enquanto os meses de fevereiro, março, abril, maio e outubro são os que menos se consome. Os meses de agosto, setembro, novembro e dezembro ficam próximo da média anual que é 2.382.145MW/mês.

Como mostra a Figura 3, os meses de janeiro, junho e julho tem um consumo de quase 10% mais do que a média anual, enquanto fevereiro, março, abril e maio tem um consumo de aproximadamente 7,5% menos do que a média anual.

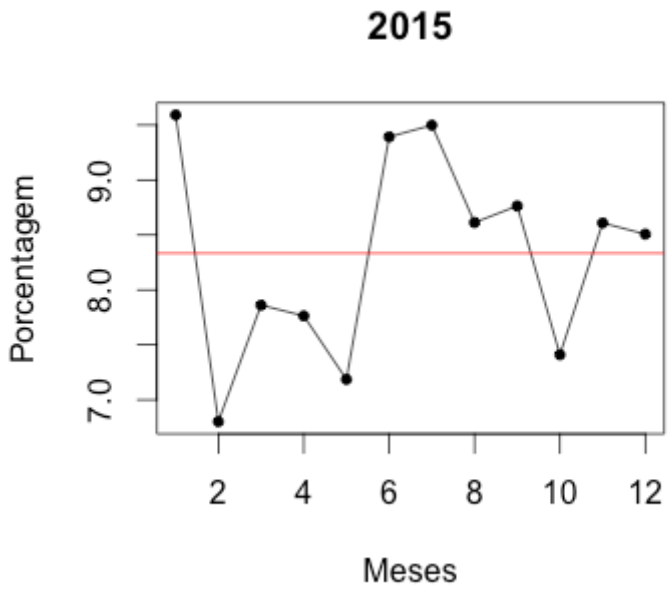

Figura 3: Consumo Mensal e média anual

3.1.1 Análise por Agente Comercializador

Os dados obtidos sobre a venda de cada Agente são mostrados na Tabela 3

	Agente	Venda 2015 em MW
1	AUDAC02	712.354,4
2	AUDP02	6.712,1
3	EDPC2	19.252.396,3
4	EDPSVD1	20.136.193,1
5	EDPUC2	6.746.958,8
6	EGEDC02	763,5
7	EGLEC2	338.419,2
8	ENATC02	27.001,5
9	ENDEC2	0
10	ENDPC2	7.638.550,7
11	ENFOC02	134.006
12	FORTIC2	1.363.542,2
13	GALPWC2	3.492.339,2
14	GNCOC02	1.166.767,7
15	GNSEC02	948.739,2
16	GOLDC02	178.774,5
17	IBCOMC2	0
18	ICLIC02	5.540.749,4
19	IGESC02	0
20	IGESC2	1.147.873,1
21	NEXUC02	0
22	ELUSC02	0
23	EYGAC02	0
24	HENSC02	7.459,7
25	LUZBC02	3.603,4
26	PHENC02	14.922,1
27	VOLTC02	0
	TOTAL	68.858.126,1

Tabela 3: Análise por Agente Comercializador

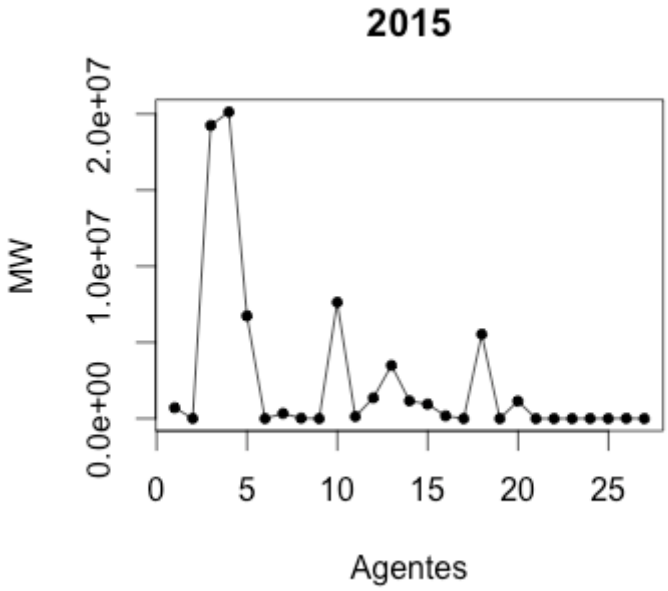

Figura 4: Venda de energia por Agente Comercializador

Como é possível verificar na Tabela 3, os dois Agentes que mais venderam em 2015 foram EDPC2 e EDPSVD1 com aproximadamente 20 milhões de MW. Em seguida temos EDPUC2, ENDPC2 e ICLIC02 que venderam entre 5 e 8 milhões de MW.

3.1.1.1 Comparação entre 2013, 2014 e 2015 .

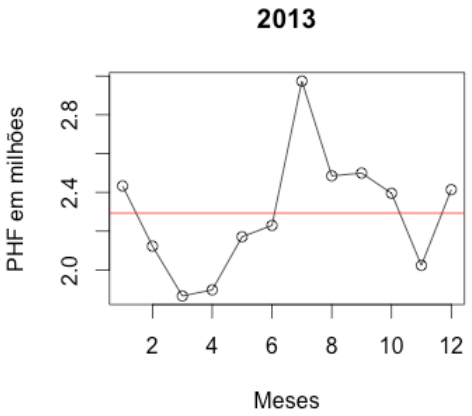

Figura 5: PHF 2014

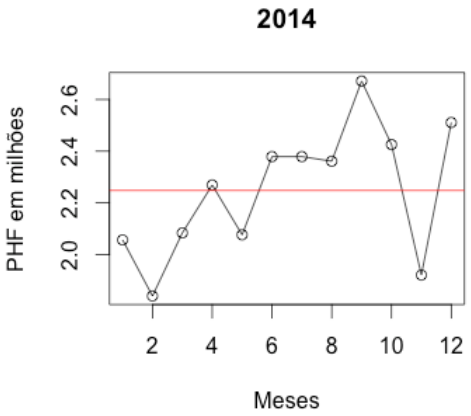

Figura 6: PHF 2015

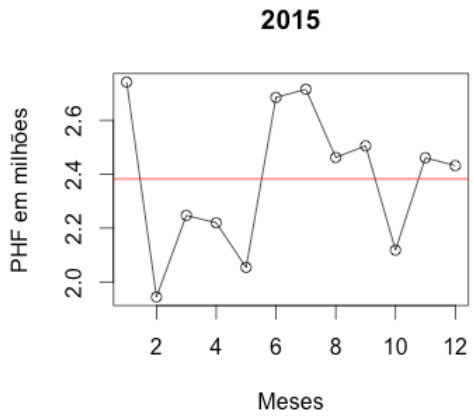

Figura 7: PHF 2016

Mês	2013	2014	2015
Janeiro	2.433.224	2.056.875	2.741.924
Fevereiro	2.122.687	1.840.074	1.944.388
Março	1.865.718	2.084.014	2.247.228
Abril	1.897.105	2.269.280	2.219.374
Maio	2.171.130	2.076.747	2.054.180
Junho	2.229.580	2.379.115	2.684.959
Julho	2.973.940	2.379.115	2.715.212
Agosto	2.484.998	2.361.240	2.462.078
Setembro	2.499.382	2.671.244	2.505.219
Outubro	2.395.350	2.425.763	2.118.323
Novembro	2.023.459	1.921.043	2.460.992
Dezembro	2.414.104	2.510.919	2.431.862
TOTAL	27.510.677	26.975.429	28.585.740

Tabela 4: Comparação 2014, 2015 e 2016

3.2 Machine Learning

O desenvolvimento da metodologia de previsão probabilística será realizada com Machine Learning, onde consiste em introduzir dados para que o algoritmo "aprenda" o comportamento das variáveis e obtenha o reconhecimento dos padrões. [2][6]

O Machine Learning terá duas etapas:

1: Treino

2: Teste/Previsão

Na primeira etapa serão utilizados os dados de eventos passados, já incluindo o Target que são os Desvios.

Na segunda etapa serão utilizadas as variáveis explicativas e o output sera o Target (Desvios).

3.2.1 Criação da Base de Dados

Para poder ser iniciado o processo de desenvolvimento da metodologia de previsão, primeiro é necessário criar a base de dados com as variáveis.

As variáveis serão:

Ano, mês, dia do mês, dia da semana, hora, unidade comercializadora, oferta (PHF) e desvio (Target) (variáveis obtidas em [12]).

Previsão de geração de energia eólica, previsão de geração de energia solar e previsão de carga total para Portugal (variáveis obtidas em [13]).

ANO	MES	DIA_MES	DIA_SEMANA	HORA	UPROG	PHF	CARGA_PREV	CARGA_REAL	DESVIO
2015	1	1	5	1	AUDAC02	-38.2	5228	5606	-0.910
2015	1	1	5	2	AUDAC02	-33.0	5010	5341	-6.556
2015	1	1	5	3	AUDAC02	-31.3	4820	5124	-8.305
2015	1	1	5	4	AUDAC02	-30.6	4521	4771	-7.420
2015	1	1	5	5	AUDAC02	-30.6	4250	4444	-6.641
2015	1	1	5	6	AUDAC02	-29.8	4083	4235	-5.915
2015	1	1	5	7	AUDAC02	-30.6	3981	4118	-4.752
2015	1	1	5	8	AUDAC02	-32.0	3940	4074	-4.150
2015	1	1	5	9	AUDAC02	-34.8	3797	3921	-0.892
2015	1	1	5	10	AUDAC02	-42.2	3769	3884	4.692
2015	1	1	5	11	AUDAC02	-44.8	4045	4235	6.778
...	...	...	...	...	...	...	...	...	...

Tabela 5: Base de Dados de 2015 para do agente AUDAC02

Como é possível observar na Tabela 5, para cada agente serão 10 variáveis dividias por hora do dia. Portanto para cada ano serão 24x365 = 8.760 dados por agente.

3.2.2 Missing Values

Foi observado a presença de *Missing Values* em diferentes variáveis.

Esses valores faltantes devem ser levados em consideração caso o valor seja significativamente suficiente para comprometer a confiabilidade da base de dados.

As tabelas 7 e 9 mostram o resumo de *missing values* para cada Unidade.

3.3 Base de Dados de 2015

Com os dados obtidos em [12] e [13] foi criada a Base de Dados de 2015 das 27 Unidades Comercializadoras.

As variáveis de entrada foram: Ano, Mês, Dia do Mês, Dia da Semana, Hora, Unidade, PHF, Previsão de Carga Total, Previsão de Geração de Energia Eólica, Previsão de Geração de Energia Solar e Desvio.

ANO	MES	DIA_MES	DIA_SEMANA	HORA	UPROG	PHF	CARGA_PREV	EOLICA_PREV	SOLAR_PREV	DESVIO
...	...	...	...	...	...	...	...	...	...	...
2015	3	12	5	9	AUDAC02	-84.9	5307	371	32	-5.291
2015	3	12	5	10	AUDAC02	-122.7	6111	335	136	-5.072
2015	3	12	5	11	AUDAC02	-132.3	6439	269	224	-8.013
2015	3	12	5	12	AUDAC02	-132.3	6366	246	287	-9.915
2015	3	12	5	13	AUDAC02	-134.6	6368	283	324	-10.426
2015	3	12	5	14	AUDAC02	-115.5	6238	377	340	-9.684
2015	3	12	5	15	AUDAC02	-116.0	6194	547	335	-11.879
2015	3	12	5	16	AUDAC02	-128.0	6250	772	307	-12.691
2015	3	12	5	17	AUDAC02	-124.7	6194	983	256	-11.401
2015	3	12	5	18	AUDAC02	-108.3	6092	1148	183	-10.584
2015	3	12	5	19	AUDAC02	-93.0	6009	1269	87	-10.357
...	...	...	...	...	...	...	...	...	...	...

Tabela 6: Base de Dados de 2015

A tabela 6 mostra uma parte da Base de Dados de 2015 para a Unidade AUDAC02.

3.3.1 Resumo de cada Unidade Comercializadora em 2015

Ao longo do processo de criação da base de dados, muitos problemas foram sendo encontrados: Missing Values, dados incompletos e falta de dados. Em algumas Unidades faltam muitos dados de PHF e Desvios e em alguns casos não há nenhum dado.

	Unidade	Resumo de Dados
1	AUDAC02	Dados Completos
2	AUDPC02	Não há PHF de 01-01 até 15-05. Possui 15 Missing Values de Desvios.
3	EDPC2	Dados Completos
4	EDPSVD1	Dados Completos
5	EDPUC2	Dados Completos
6	EGEDC02	Não há Desvios de 01-01 até 10-01
7	EGLEC2	2 Missing Values de Desvio
8	ENATC02	13 Missing Values de Desvio
9	ENDEC2	Não há PHF nem Desvios em 2015
10	ENDPC2	Dados Completos
11	ENFOC02	Dados Completos
12	FORTIC2	Dados Completos
13	GALPWC2	Dados Completos
14	GNCOC02	Dados Completos
15	GNSEC02	Dados Completos
16	GOLDC02	Dados Completos
17	HENSC02	Faltam 648 PHF e 756 Desvios
16	IBCOMC2	Não há PHF nem Desvios em 2015
19	ICLIC02	Dados Completos
20	IGESC02	Não há PHF nem desvios em 2015
21	IGESC2	1 Missing Value de Desvio
22	LUZBC02	Faltam 624 PHF e 536 Desvios
23	NEXUC02	Não há PHF nem Desvios em 2015
24	ELUSC02	Não há PHF e 3360 Missing Values de Desvios
25	PHENC02	Faltam 2111 PHF e 2544 Desvios
26	VOLTC02	Não há PHF nem Desvios em 2015
27	EYGAC02	Não há PHF nem Desvios em 2015

Tabela 7: Dados das Unidades em 2015

3.4 Base de Dados de 2016

Como o objetivo é a criação de uma base de dados com 18 meses de dados, para 2016 foi utilizado os valores de 01/jan a 30/06.

ANO	MES	DIA_MES	DIA_SEMANA	HORA	UPROG	PHF	CARGA_PREV	EOLICA_PREV	SOLAR_PREV	DESVIO
...	...	...	...	...	...	...	...	...	...	...
2016	5	08	1	21	AUDAC02	-45.2	3658	36	4668	-0.183
2016	5	08	1	22	AUDAC02	-45.4	3622	4	5108	0.447
2016	5	08	1	23	AUDAC02	-44.9	3647	0	5344	0.988
2016	5	08	1	24	AUDAC02	-45.8	3654	0	5210	2.029
2016	5	09	2	1	AUDAC02	-42.9	3426	0	4787	1.600
2016	5	09	2	2	AUDAC02	-45.6	3417	0	4408	2.830
2016	5	09	2	3	AUDAC02	-45.4	3352	0	4157	3.408
2016	5	09	2	4	AUDAC02	-45.1	3263	0	4039	1.583
2016	5	09	2	5	AUDAC02	-45.8	3213	0	3984	2.647
2016	5	09	2	6	AUDAC02	-44.4	3211	0	3988	1.004
2016	5	09	2	7	AUDAC02	-47.3	3176	0	4030	1.804
2016	5	09	2	8	AUDAC02	-53.0	3156	8	4156	0.951
2016	5	09	2	9	AUDAC02	-65.4	3096	51	4772	-1.381
...	...	...	...	...	...	...	...	...	...	...

Tabela 8: Base de Dados de 2016

3.4.1 Resumo de cada Unidade Comercializadora em 2016

Assim como nos dados de 2015, houveram Missing Values, dados incompletos, falta de dados, em algumas Unidades faltam muitos dados de PHF e Desvios e em alguns casos não há nenhum dado.

	Unidade	Resumo de Dados
1	AUDAC02	Faltam 2 Desvios e 1 PHF
2	AUDPC02	Faltam 4 Desvios e 1 PHF
3	EDPC2	Falta 1 Desvio e 1 PHF
4	EDPSVD1	Falta 1 Desvio e 1 PHF
5	EDPUC2	Falta 1 Desvio e 1 PHF
6	EGEDC02	Faltam 8 Desvios e 1 PHF
7	EGLEC2	Faltam 2 Desvios e 1 PHF
8	ELUSC02	Falta 1 Desvio e 1 PHF
9	ENATC02	Faltam 8 Desvios e 1 PHF
10	ENDEC2	Não há PHF nem Desvios
11	ENDPC2	Falta 1 Desvio e 1 PHF
12	ENFOC02	Faltam 3 Desvios e 1 PHF
13	EYGAC02	Faltam 1326 Desvios e 1 PHF
14	FORTIC2	Falta 1 Desvio e 1 PHF
15	GALPWC2	Falta 1 Desvio e 1 PHF
16	GNCOC02	Falta 1 Desvio e 1 PHF
17	GNSEC02	Falta 1 Desvio e 1 PHF
18	GOLDC02	Faltam 2 Desvios e 1 PHF
19	HENSC02	Faltam 11 Desvios e 1 PHF
20	IBCOMC2	Não há PHF nem Desvios
21	ICLIC02	Falta 1 Desvio e 1 PHF
22	IGESC02	Não há PHF nem Desvios
23	IGESC26	Falta 1 Desvio e 1 PHF
24	LOGIC026	PHF a partir de 27/jan e Desvios a partir de 05/03 – Faltam 67 Desvios
25	LUZBC02	Faltam 7 Desvios e 1 PHF
26	NEXUC02	Não há PHF nem Desvios
27	PHENC02	Falta 1 Desvio e 1 PHF
28	VOLTC02	Não há PHF nem Desvios
29	ELERC02	PHF a partir de 22/mar e Desvios a partir de 01/abril
30	ECOCC02	PHF a partir de 29/abril e Desvios a partir de 13/maio
31	ROLEC02	Não há PHF nem Desvios

Tabela 9: Dados das Unidades em 2016

3.5 Base de Dados Total

A Base de Dados Total é composta pelas seguintes variáveis:

- ANO
- MES
- DIA_ MÊS
- DIA_SEMANA
- HORA
- UPROG
- PHF
- PREV_EOLICA
- PREV_SOLAR
- PREV_CARGA
- DESVIO

Os dados foram recolhidos entre 01-01-2015 e 30-06-2016.

Ao total foram 31 Unidades Comercializadoras, sendo 10 variáveis (para cada variável há 13.129 dados).

Ao total foram identificados 178 Missing Values.

Foi criada uma Base de Dados para cada Unidade Comercializadora e uma Base de Dados com os valores totais de PHF e Desvios. Para a Base de Dados com os valores totais, foi retirada a Unidade Agente EDPSVD1 (única que vende solar e eólica).

4 Implementação Prática

Como o objetivo é realizar previsões tanto para os valores totais, ou seja, a soma dos valores de todas as unidades de energia elétrica e também prever os desvios para uma única unidade, para a realização da implementação prática foram utilizadas duas base de dados: DB_TOTAL que possui os valores totais e DB_AUDAC02 que possui os valores da unidade AUDAC02.

A unidade AUDAC02 foi escolhida por possuir uma base de dados mais completa e com menos *missing values*.

4.1 Análise das Variáveis Explicativas

Ao total são 9 variáveis explicativas, sendo elas: Ano, Mês, Dia do Mês, Dia da Semana, Hora, PHF, Previsão de Geração de Energia Eólica, Previsão de Geração de Energia Solar, Previsão de Carga.

Para uma análise preliminar foi utilizado a Base de Dados Total, onde constam os valores totais de PHF e Desvio. Nessa análise os Missing Values serão ignorados, assim a Base de Dados será: DB_TOTAL_S_NA

4.1.1 Variáveis Categóricas

Como a variável Ano possui apenas dois levels (2015 e 2016) trataremos ela como uma variável categórica. As variáveis Mês, Dia do Mês, Dia da Semana e Hora não são categóricas, porém por hora não serão analisadas como contínuas.

4.1.2 Variáveis Contínuas

4.1.2.1 PHF

A Variável PHF (Programa Horário Final) é a previsão de produção/venda de energia elétrica em MWh. Quando o valor é negativo, indica venda, e quando o valor é positivo indica produção de energia elétrica. Apenas a Unidade Comercializadora EDPSVD1 possui valores positivos. Por esse motivo, essa Unidade foi excluída da Base de Dados Total (BD_TOTAL).

Conforme os gráficos abaixo, é possível notar que a variável tende a uma distribuição Normal, e possui outlier que deverá ser estudado e se for necessário [30], retirado da base de dados.

Summary:

Min.	1st Qu	Median	Mean	3rd Qu	Max.
-7031	-5418	-4809	-4790	-4040	-2963

Tabela 10: Summary PHF

Desvio Padrão: 850.4208

Correlação com o Target (DESVIO):

Coeficiente de Pearson: -0.008259984

Kendall: -0.006053711

Spearman: -0.00888764

Conclusão: Pouca Correlação.

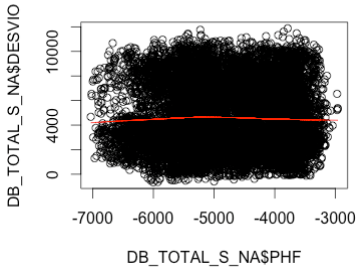

Figura 8: Gráfico PHF com a Resposta

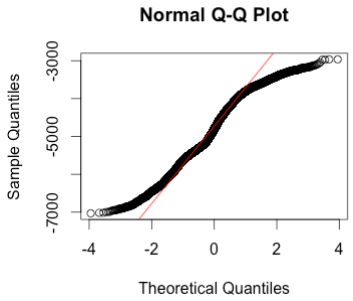

Figura 9: Gráfico Q-Q Norm da variável PHF

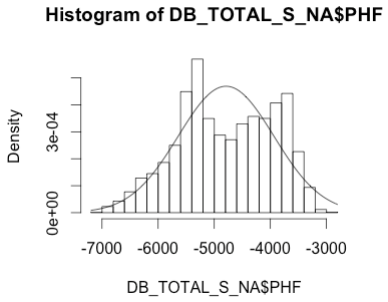

Figura 10: Histograma da variável PHF

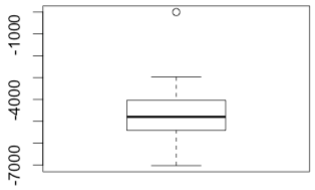

Figura 11: Boxplot da variável PHF (Verificado outlier)

4.1.2.2 Previsão de Geração de Energia Eólica

Summary:

Min.	1st Qu	Median	Mean	3rd Qu	Max.
0	529	1077	1344	1936	4252

Tabela 11: Summary Previsão de Geração de Energia Eólica

Desvio Padrão: 1016.01

Correlação com o Target (DESVIO):

Coeficiente de Pearson: -0.1218135

Kendall: -0.08344155

Spearman: -0.1243758

Conclusão: Pouca Correlação.

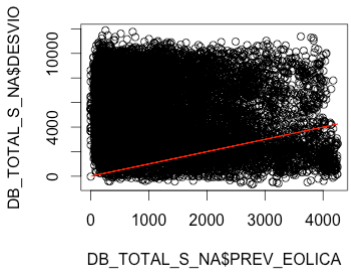

Figura 12: Gráfico Prev. Eólica com a Resposta

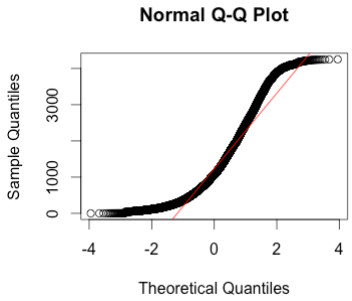

Figura 13: Gráfico Q-Q Norm da variável Prev. Eólica

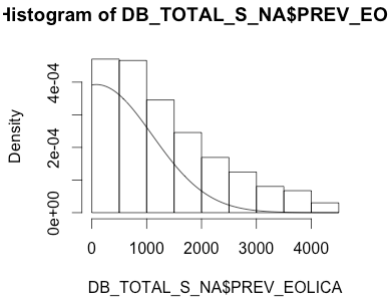

Figura 14: Histograma da variável Prev. Eólica

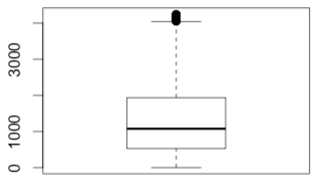

Figura 15: Boxplot da variável Prev. Eólica

4.1.2.3 Previsão de Geração de Energia Solar

Summary:

Min.	1st Qu	Median	Mean	3rd Qu	Max.
0.00	0.00	3.00	84.52	169.00	523.00

Tabela 12: Summary Previsão de Geração de Energia Solar

Desvio Padrão: 112.7775

Correlação com o Target (DESVIO):

Coeficiente de Pearson: -0.09976527

Kendall: -0.07505332

Spearman: -0.1056223

Conclusão: Pouca Correlação.

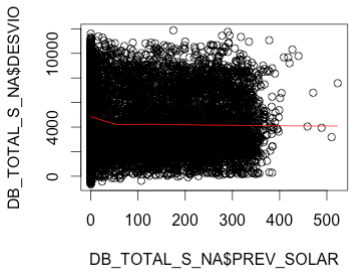

Figura 16: Gráfico Prev. Solar com a Resposta

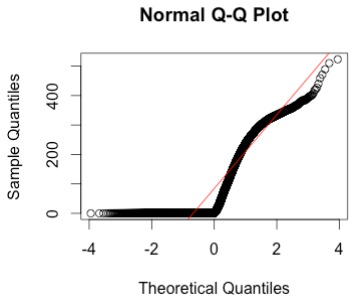

Figura 17: Gráfico Q-Q Norm da variável Prev. Solar

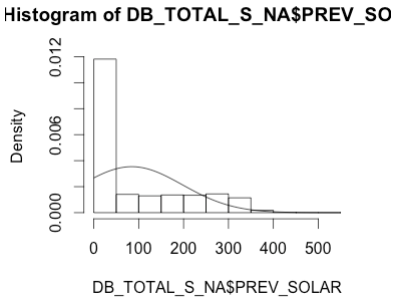

Figura 18: Histograma da variável Prev. Eólica

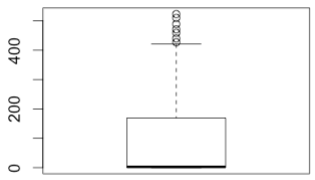

Figura 19: Boxplot da variável Prev. Solar

4.1.2.4 Previsão de Carga

Summary:

Min.	1st Qu	Median	Mean	3rd Qu	Max.
3462	4769	5633	5624	6344	8475

Tabela 13: Summary Previsão de Carga

Desvio Padrão: 973.6113

Correlação com o Target (DESVIO):

Coeficiente de Pearson: -0.01479269

Kendall: -0.008017729

Spearman: -0.01284082

Conclusão: Pouca Correlação.

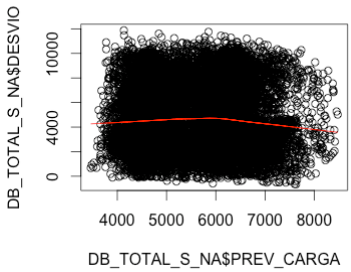

Figura 20: Gráfico Prev. Carga com a Resposta

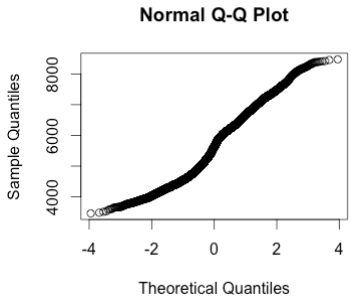

Figura 21: Gráfico Q-Q Norm da variável Prev. Carga

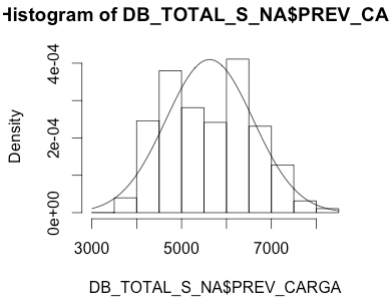

Figura 22: Histograma da variável Prev. Carga

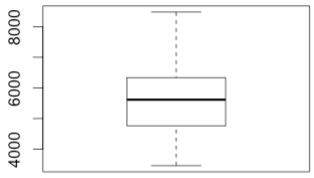

Figura 23: Boxplot da variável Prev. Carga

4.1.3 Correlação entre as Variáveis

Conforme Figura 24 é possível verificar que há uma forte correlação inversa entre as variáveis PREV_CARGA e PHF:

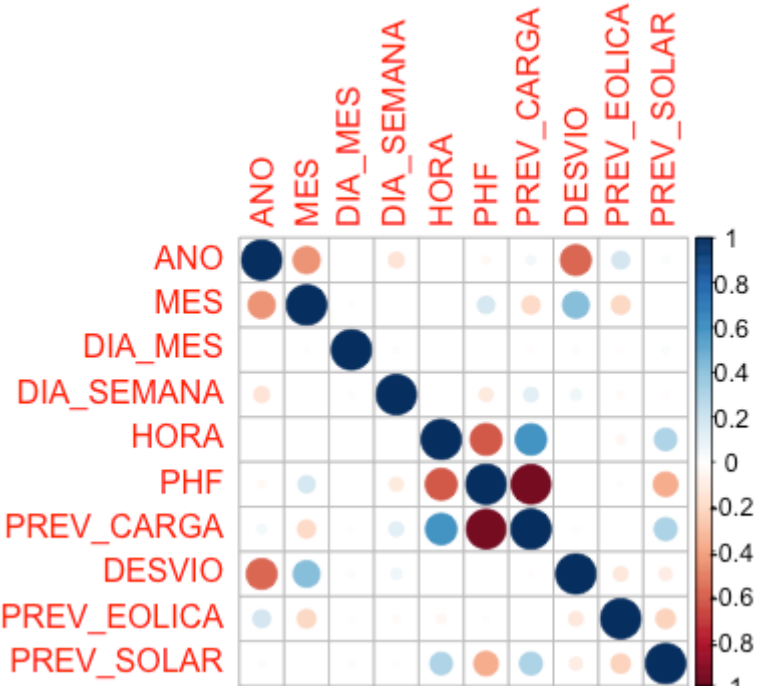

Figura 24: Correlação entre as Variáveis

Correlação entre PREV_CARGA e PHF:

Coeficiente de Pearson: -0.9585517

Kendall: -0.824168

Spearman: -0.9597139

Conclusão: Forte Correlação.

Deverá ser estudado se essa correlação não prejudica o modelo. Caso positivo, deverá ser excluída a menos significativa.

4.2 Análise dos Modelos da DB_TOTAL

Antes da aplicação dos Métodos de Previsão, é necessário analisar os modelos da base de dados. Ou seja, verificar quais as variáveis mais significativas e analisar se as menos significativas devem ser retiradas do modelo ou transformadas. É também necessário verificar se a presença de *Missing Values* e *Outliers* prejudicam o modelo e devem ser retiradas.

Foi então realizado um primeiro teste de previsão com Regressão Linear para observar como o modelo completo se comporta.

Com a utilização da função *lm()* contemplando todas as variáveis, foram obtidos os seguintes dados:

Min	1Q	Median	3Q	Max
-5868.4	-1474.5	-165.5	1194.3	9386.6

Tabela 14: Residuals

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.431e+06	9.109e+04	59.622	<2e-16 ***
DB_TOTAL$ANO	-2.694e+03	4.519e+01	-59.615	<2e-16 ***
DB_TOTAL$MES	1.922e+02	6.512e+00	29.518	<2e-16 ***
DB_TOTAL$DIA_MES	7.871e+00	2.117e+00	3.718	0.000202 ***
DB_TOTAL$DIA_SEMANA	-2.598e+01	8.883e+00	-2924	0.003460 **
DB_TOTAL$HORA	-1.801e+01	3.481e+00	-5.175	2.32e-07 ***
DB_TOTAL$PHF	-8.130e-01	7.997e.02	-10.166	<2e-16 ***
DB_TOTAL$PREV_EOLICA	-3.613e.02	1.935e-02	-1.867	0.061929 .
DB_TOTAL$PREV_SOLAR	-3.214e+00	1.849e-01	-17.382	<2e-16 ***
DB_TOTAL$PREV_CARGA	-3.245e-01	6.814e-02	-4763	1.93e-06 ***

Tabela 15: Coefficients

Sendo que os códigos de significância são: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 " 1

É possível verificar que a variável menos significativa é PREV_EOLICA.

Também foi verificado que o p-value para esse modelo é baixo (< 2.2e-16), ou seja, rejeitamos a hipótese nula de que não existe relação entre os fenômenos medidos.

O R Ajustado é de 0.3918, não é um bom valor, porém se tratando de um modelo completo, onde não foi retirado *outliers* das variáveis, não foi retirada nenhuma variável menos significativa e não foi realizado nenhuma transformação para melhorar o modelo, podemos continuar com as análises dessa regressão.

Para verificar a eficácia do modelo completo, foi realizado um teste real com as seguintes variáveis: ANO: 2016

MES: 7

DIA_MES: 1

DIA_SEMANA: 6

HORA: 15

PHF: -3074,1

PREV_EOLICA: 833

PREV_SOLAR: 323

PREV_CARGA: 6634

c(1, 2016, 7, 1, 6, 15, -3074.1, 833, 323, 6634)%*%modelo$coefficients

O resultado foi 412.4555 e o valor real é 318.775. Portanto um erro de 22.71287%.

A Variável PREV_EOLICA é a menos significativa para o modelo, então foi retirada essa variável e criado o modelo_2:

Min	1Q	Median	3Q	Max
-5840.1	-1474.5	-163.7	1188.4	9431.2

Tabela 16: Residuals

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.451e+06	9.050e+04	60.225	<2e-16 ***
DB_TOTAL$ANO	-2.704e+03	4.489e+01	-60.220	<2e-16 ***
DB_TOTAL$MES	1.938e+02	6.457e+00	30.018	<2e-16 ***
DB_TOTAL$DIA_MES	7.890e+00	2.118e+00	3.726	0.000195 ***
DB_TOTAL$DIA_SEMANA	-2.563e+01	8.882e+00	-2885	0.003915 **
DB_TOTAL$HORA	-1.799e+01	3.481e+00	-5.168	2.41e-07 ***
DB_TOTAL$PHF	-8.107e-01	7.997e-02	-10.137	<2e-16 ***
DB_TOTAL$PREV_SOLAR	-3.135e+00	1.799e-01	-17.419	<2e-16 ***
DB_TOTAL$PREV_CARGA	-3.244e-01	6.814e-02	-4e760	1.96e-06 ***

Tabela 17: Coefficients

O modelo continua com p-value baixo ($< 2.2e\text{-}16$) e o R Ajustado quase não é alterado (0.3916).

Para testar se a qualidade de ajustamento dos modelos é igual, utilizaremos ANOVA com a função *anova()*.

Res. DF	RSS	DF	Sum of Sq	F	Pr(>F)
12686	5.5525e+10				
12687	5.5540e+10	-1	-15255896	3.4856	0.06193

Tabela 18: ANOVA

O resultado foi um p-value acima de 0.05 (0.06193), portanto não rejeitamos a hipótese nula e mantemos o modelo original com a variável PREV_EOLICA.

4.3 Análise dos Modelos da DB_AUDAC02

O mesmo procedimento realizado com a DB_TOTAL foi feito com a DB_AUDAC02.

Min	1Q	Median	3Q	Max
-33.415	-2.322	0.106	2.387	27.359

Tabela 19: Residuals

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.525e+00	3.805e-01	9.266	< 2e-16 ***
DB_TOTAL$MES	1.352e-01	1.482e-02	9.123	< 2e-16 ***
DB_TOTAL$DIA_MES	-1.344e-02	5.614e-03	-2.394	0.016675 *
DB_TOTAL$DIA_SEMANA	-8.425e-02	2.500e-02	-3.370	0.000754 ***
DB_TOTAL$HORA	1.793e-02	9.318e-03	1.924	0.054417 .
DB_TOTAL$PHF	9.707e-01	2.500e-03	388.347	< 2e-16 ***
DB_TOTAL$PREV_EOLICA	-1.921e-04	5.320e-05	-3.611	0.000306 ***
DB_TOTAL$PREV_SOLAR	-6.610e-03	5.532e-04	-11.949	< 2e-16 ***
DB_TOTAL$PREV_CARGA	-1.703e-03	8.429e-05	-20.208	< 2e-16 ***

Tabela 20: Coefficients

O p-value para esse modelo é baixo (< 2.2e-16), ou seja, rejeitamos a hipótese nula de que não existe relação entre os fenômenos medidos.

O R Ajustado é de 0.9782, um valor extremamente bom que indica que o modelo é eficaz para previsão.

Os mesmos testes realizados com a DB_TOTAL, utilizando a ANOVA, foram realizados com a DB_AUDAC02, e novamente não se mostrou necessário retirar as variáveis menos significativas.

4.4 Resultados

Para a realização da implementação prática foi utilizado a Base de Dados Completa Sem Missing Values (DB_TOTAL_S_MV) e a Base de Dados AUDAC02 Sem Missing Values (DB_AUDAC02_S_MV).

Utilizando testes com ANOVA, foi verificado a não necessidade de retirar outliers, porém, foi realizado a transformação da variável dependente para PHF + DESVIO.

A Base de Dados foi dividida em duas para cada um dos dois testes: Base de Dados de Treino (DB_TOTAL_TREINO) e (DB_AUDAC02L_TREINO) que compreende os dados de 01-01-2015 a 31-12-2015; e Base de Dados de Teste (DB_TOTAL_TESTE) e (DB_AUDAC02_TESTE) que compreende os dados de 01-01-2016 a 30-06-2016.

4.4.1 Resultados com Regressão Linear Múltipla

4.4.1.1 Base de Dados Total

Min	1Q	Median	3Q	Max
-1308.5	-324.9	-79.8	295.7	2468.0

Tabela 21: Residuals

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-3.563e+02	3.786e+01	-9.411	<2e-16 ***
DB_TOTAL$MES	1.347e+02	1.505e+00	89.462	<2e-16 ***
DB_TOTAL$DIA_MES	2.923e+00	5.688e-01	5.138	2.84e-07 ***
DB_TOTAL$DIA_SEMANA	-2.110e+00	2.516e+00	-0.839	0.401759
DB_TOTAL$HORA	-5.659e-01	9.359e-01	-0.605	0.545449
DB_TOTAL$PHF	7.091e-01	1.896-02	37.403	<2e-16 ***
DB_TOTAL$PREV_EOLICA	-1.950e-02	5.605e-03	-3.480	0.000504 ***
DB_TOTAL$PREV_SOLAR	6.946e-01	5.049e-02	13.758	<2e-16 ***
DB_TOTAL$PREV_CARGA	-2.333e-01	1.820e-02	-12.822	<2e-16 ***

Tabela 22: Coefficients

O modelo possui muitas variáveis significantes e um R-Quadrado Ajustado de 0.8412. Com o p-value $<$ 2.2e-16 rejeitamos a hipótese nula. Um resultado consideravelmente melhor comparado com o modelo total original (Tabela 15).

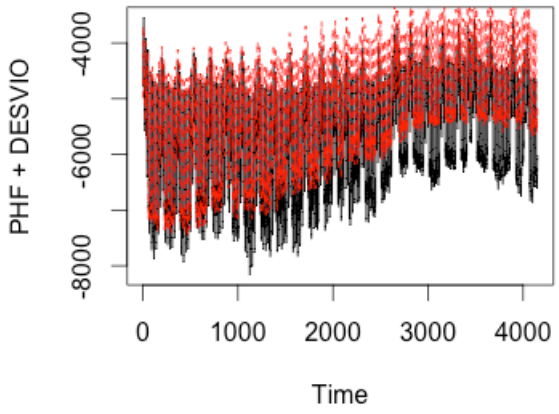

Figura 25: Valores Previstos e Valores Reais

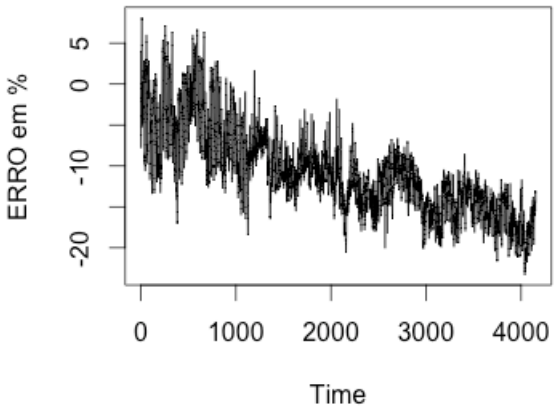

Figura 26: ERRO

A média de erro no modelo de previsão foi de 10,76303%

4.4.1.2 Base de Dados AUDAC02

Para a Base de Dados da unidade AUDAC02 foi mantido o modelo apresentado na tabela 20

onde o p-value: $<$ 2.2e-16 e o R Ajustado de 0.9782.

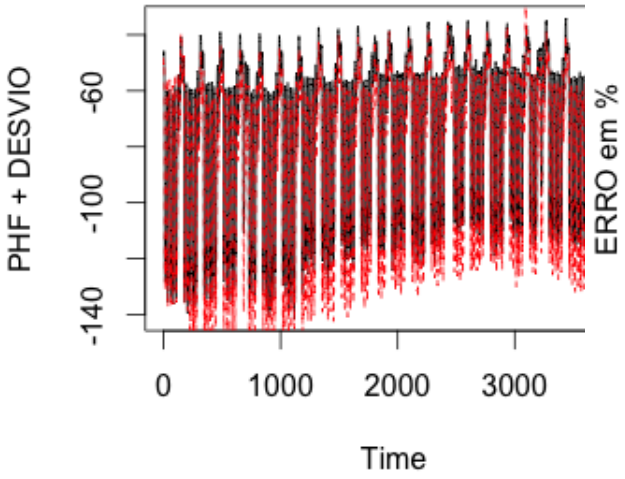

Figura 27: Valores Previstos e Valores Reais

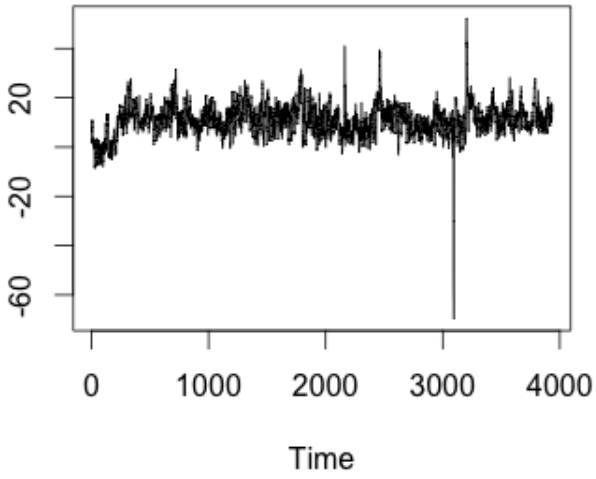

Figura 28: ERRO

A média de erro no modelo de previsão foi de 10,88228%

4.4.2 Resultados com Regressão Linear de Quantis

Foi utilizada a biblioteca *quantreg* para o método de Regressão Linear de Quantis.

4.4.2.1 Base de Dados Total

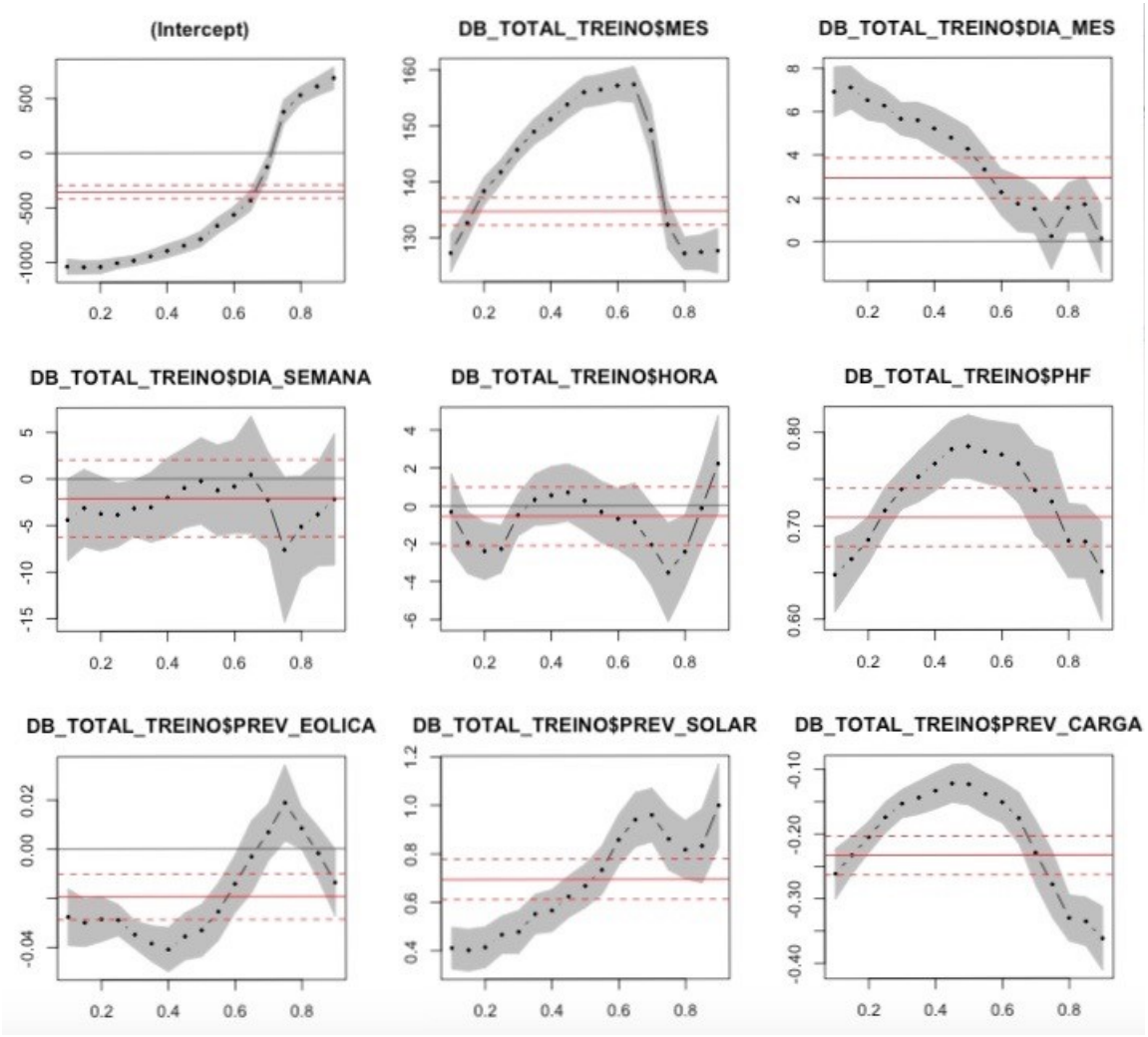

Figura 29: Variáveis Regressão Linear de Quantis

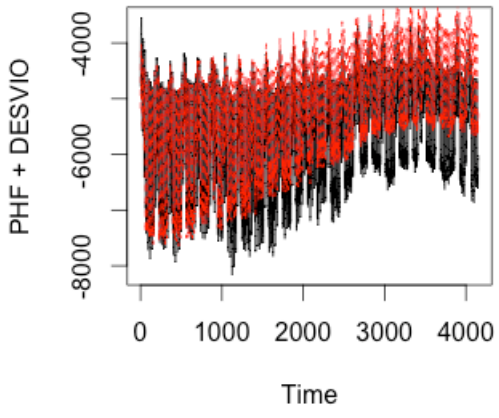

Figura 30: Regressão de Quantis – Valor Real X Previsão Q0.5

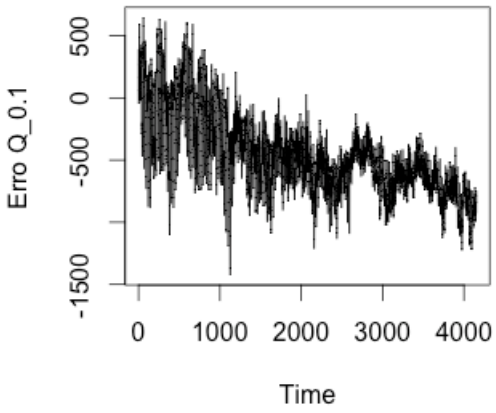

Figura 31: Regressão de Quantis – Erro Q0.5

Calibration

No capítulo 2.2.1 são apresentados os cálculos para se obter os valores de *Calibration*. O diagrama de *Calibration* é dado pela equação 2.2.6, e quanto mais próximo de zero for o desvio dos quantis estimados para os quantis nominais, melhor resultado apresenta o modelo.

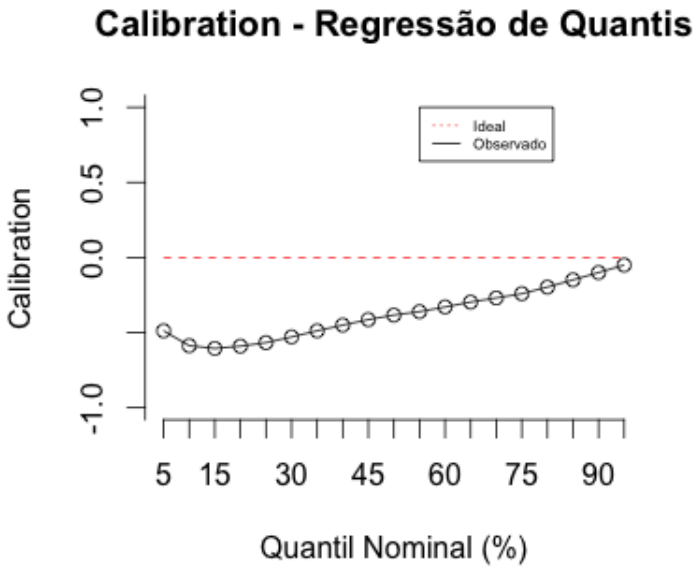

Figura 32: Regressão de Quantis – Calibration

Como é possível observar na Figura 32, a partir do quantil 15%, os valores se aproximam cada vez mais do valor ideal.

Sharpness

Com a implementação da equação 2.2.8 do capítulo 2.2.2, foi construído o diagrama de dispersão (*Sharpness*).

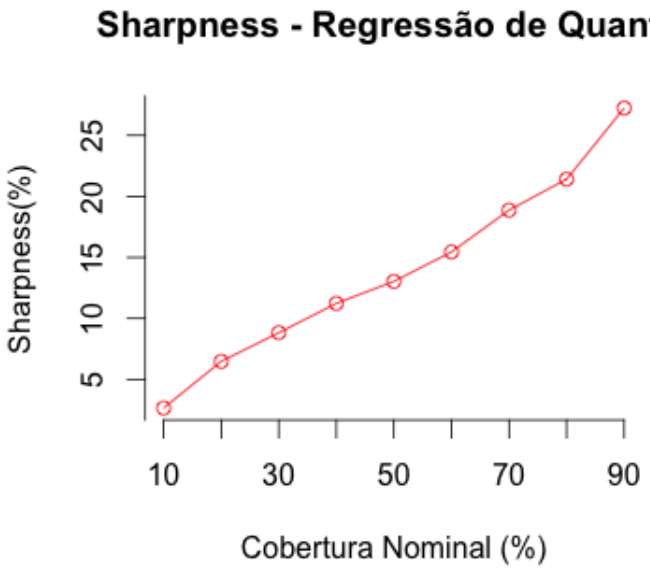

Figura 33: Regressão de Quantis – Sharpness

Como mostra a Figura 33, a dispersão entre os quantis são maiores a medida que os valores dos quantis são aumentados.

Skill Score

Para verificar a qualidade do método é utilizado o *Skill Score*, descrito no capítulo 2.2.3.

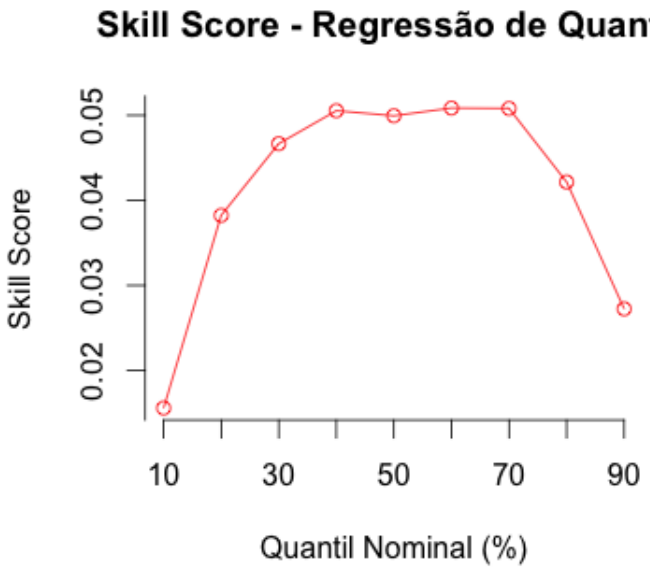

Figura 34: Regressão de Quantis – Skill Score

O *Skill Score* fornece um critério de avaliação que contém informações tanto de *Calibration* quando de *Sharpness*, podendo assim concluir a qualidade do método.

É possível observar na Figura 34 que para os quantis próximos de 50% o *Skill Score* tem pior desempenho, uma vez que quanto mais próximo de zero, melhor é o resultado. Mesmo assim, o método de Regressão de Quantis apresenta bons resultados.

Resumo dos Resultados de Desempenho para o Quantil 0.5:

MAE	RMSE	MAPE	Calibration	Shapness	Skill Score
8.0028	9.909769	8.121541	-0.3831888	0.130338	0.04994407

Tabela 23: Desempenho Regressão Linear de Quantis (Quantil 0.5)

4.4.2.2 Base de Dados AUDAC02

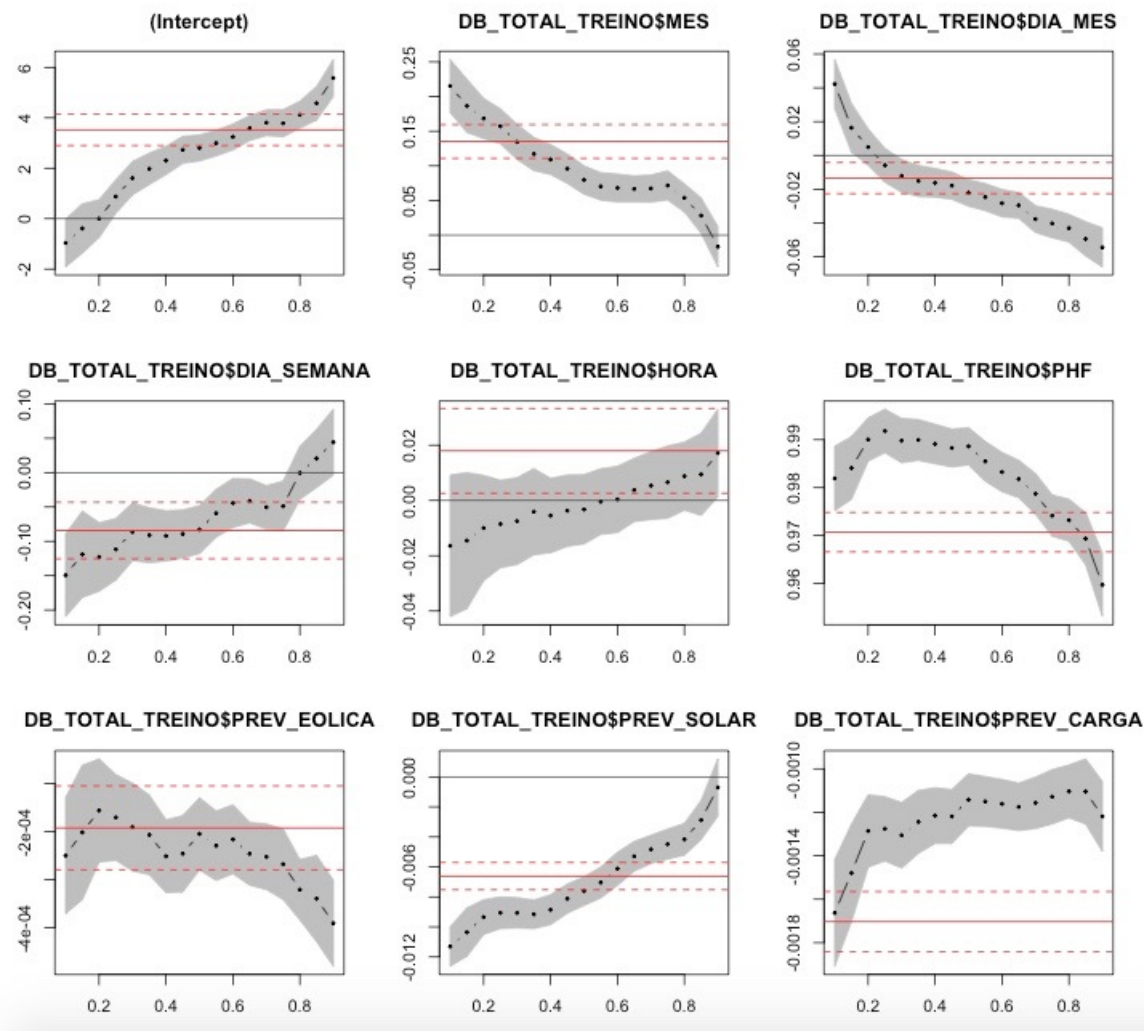

Figura 35: Variáveis Regressão Linear de Quantis

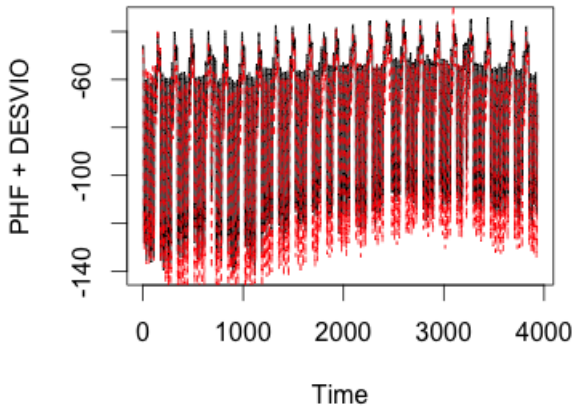

Figura 36: Regressão de Quantis – Valor Real X Previsão Q0.5

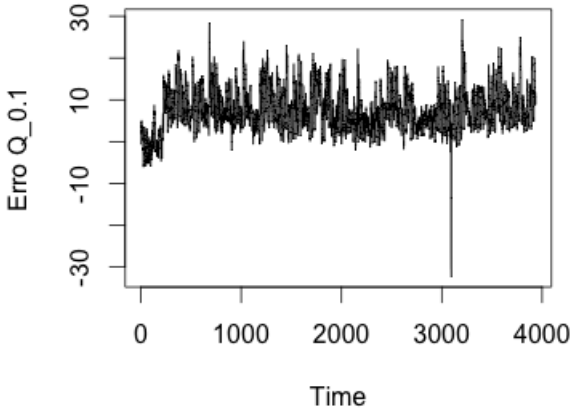

Figura 37: Regressão de Quantis – Erro Q0.5

Calibration

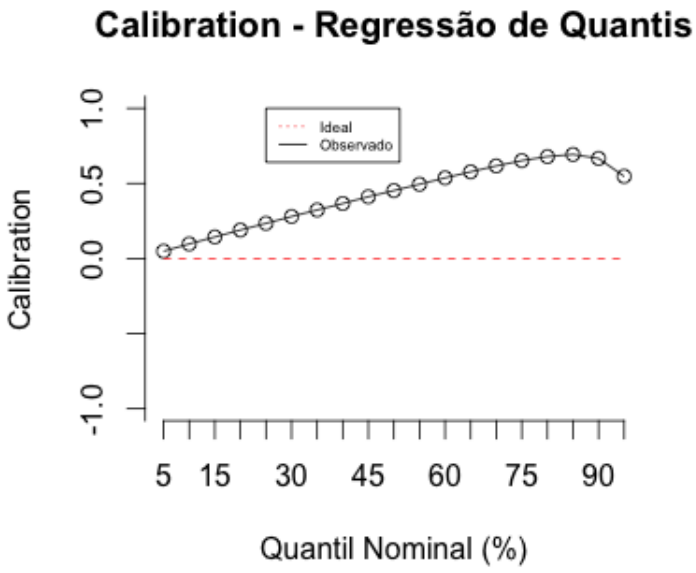

Figura 38: Regressão de Quantis – Calibration

Como é possível observar na Figura 38, até o quantil 85%, os valores se afastam cada vez mais do valor ideal, porém sempre próximo de 0.5.

Sharpness

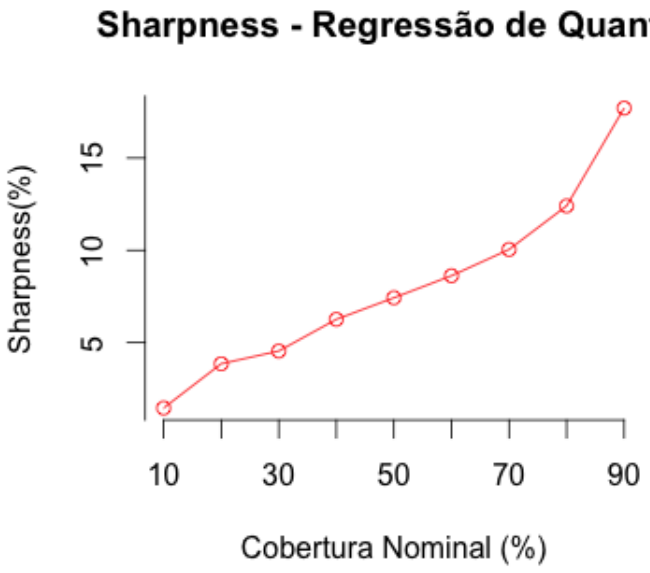

Figura 39: Regressão de Quantis – Sharpness

Como mostra a Figura 39, a dispersão entre os quantis são maiores a medida que os valores dos quantis são aumentados.

Skill Score

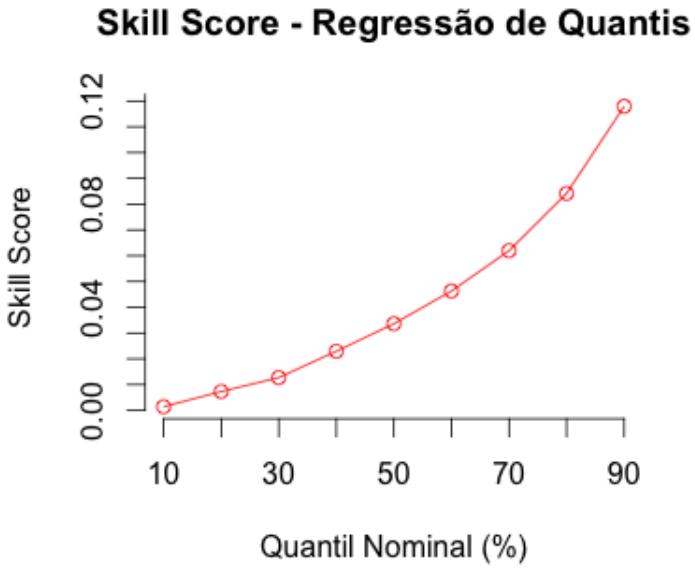

Figura 40: Regressão de Quantis – Skill Score

É possível observar na Figura 40 que o *Skill Score* tem piores resultados a medida que o valor do Quantil é aumentado.

Resumo dos Resultados de Desempenho para o Quantil 0.5:

MAE	RMSE	MAPE	Calibration	Shapness	Skill Score
10.16555	12.38017	10.23533	0.4524778	0.07428362	0.03361169

Tabela 24: Desempenho Regressão Linear de Quantis (Quantil 0.5)

4.4.3 Resultados com Quantile Regression Forests

Com a utilização do package *quantregForest* [16] foi construído o método Quantile Regression Forests, explicado no capítulo 2.1.2.1

4.4.3.1 Base de Dados Total

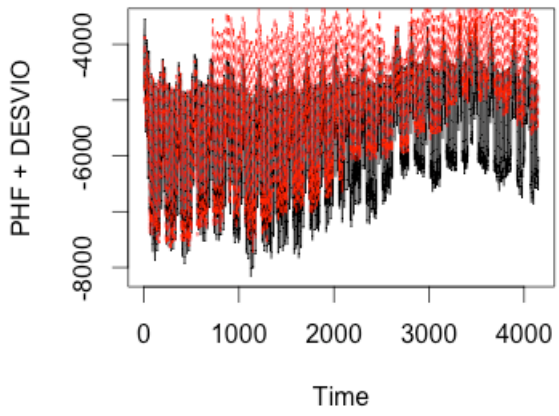

Figura 41: Quantile Regression Foreststst – Valor Real X Previsão Q0.5

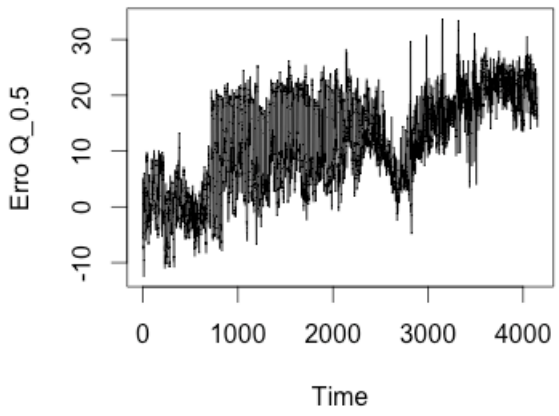

Figura 42: Quantile Regression Forests – Erro Q0.5

Calibration

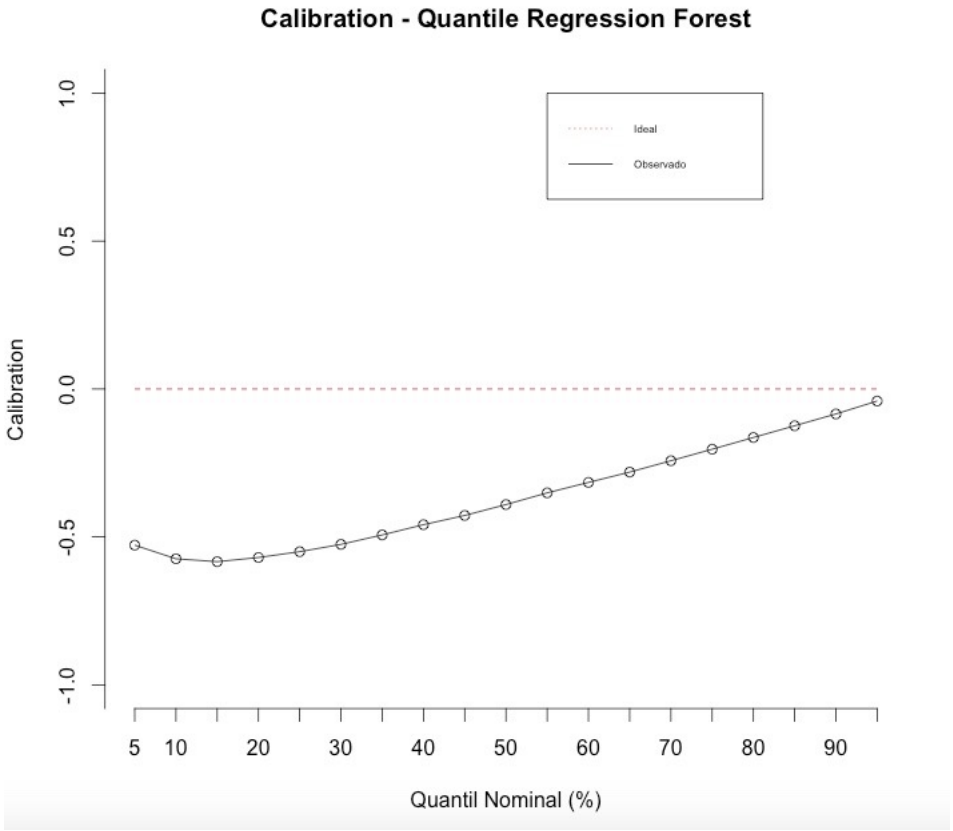

Figura 43: Quantile Regression Forests – Calibration

Como é possível observar na Figura 43, a partir do quantil 15%, os valores se aproximam cada

vez mais do valor ideal.

Sharpness

Com a implementação da equação 2.2.8 do capítulo 2.2.2, foi construído o diagrama de dispersão (*Sharpness*).

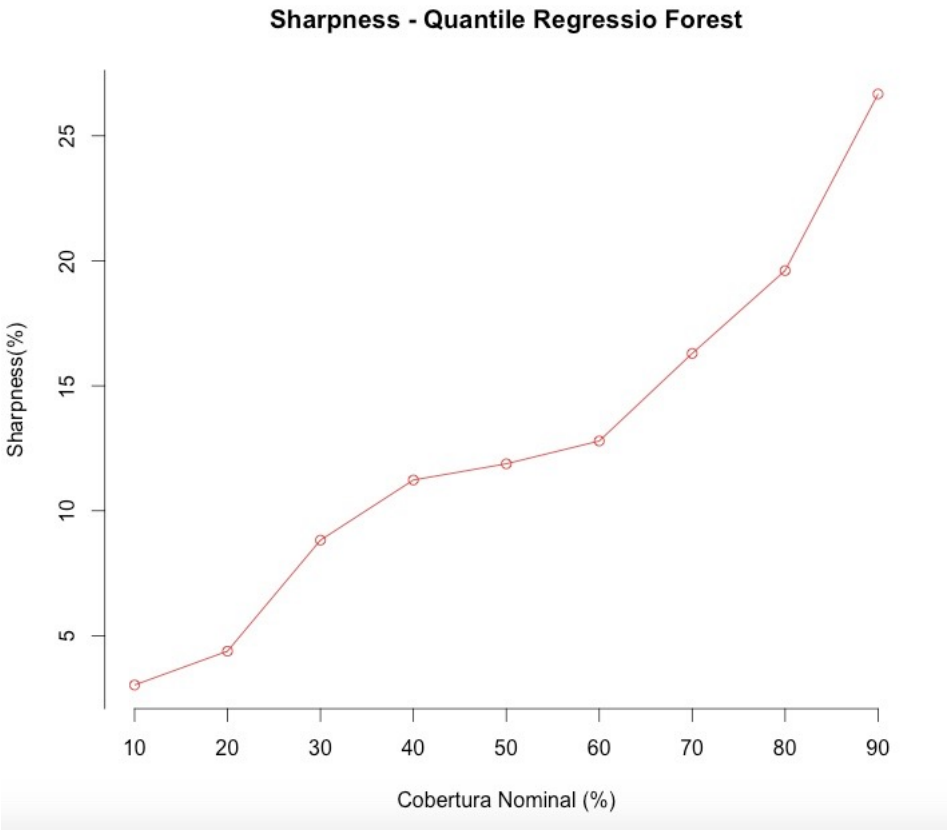

Figura 44: Quantile Regression Forests – Sharpness

Como mostra a Figura 44, a dispersão entre os quantis são maiores a medida que os valores dos quantis são aumentados.

Skill Score

Para verificar a qualidade do método é utilizado o *Skill Score*, descrito no capítulo 2.2.3.

Figura 45: Quantile Regression Forests – Skill Score

O *Skill Score* fornece um critério de avaliação que contém informações tanto de *Calibration* quando de *Sharpness*, podendo assim concluir a qualidade do método.

É possível observar na Figura 45 que para os quantis próximos de 50% o *Skill Score* tem pior desempenho, uma vez que quanto mais próximo de zero, melhor é o resultado. Mesmo assim, o método de Regressão de Quantis apresenta bons resultados.

Resumo dos Resultados de Desempenho para o Quantil 0.5:

MAE	RMSE	MAPE	Calibration	Sharpness	Skill Score
10.70286	13.52778	11.29314	-0.3906551	0.1187999	0.04640977

Tabela 25: Desempenho Quantile Regression Forest (Quantil 0.5)

4.4.3.2 Base de Dados AUDAC02

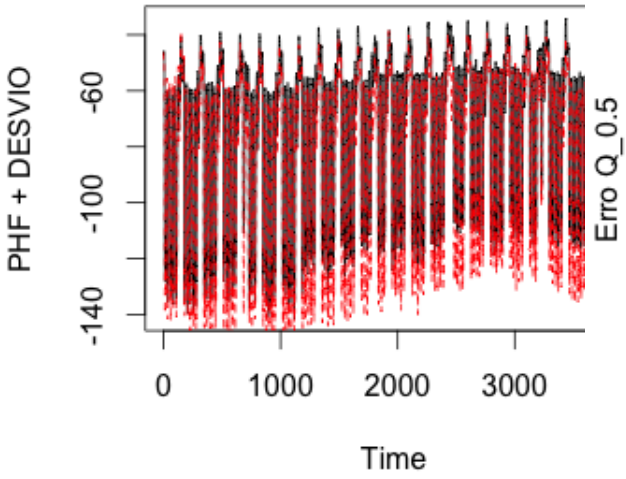

Figura 46: Quantile Regression Foreststst – Valor Real X Previsão Q0.5

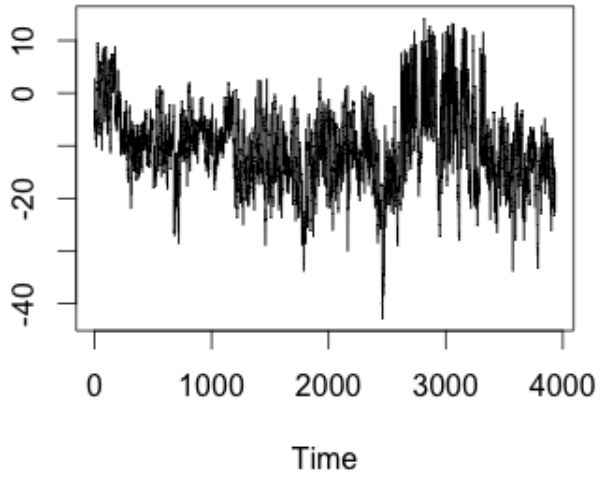

Figura 47: Quantile Regression Forests – Erro Q0.5

Calibration

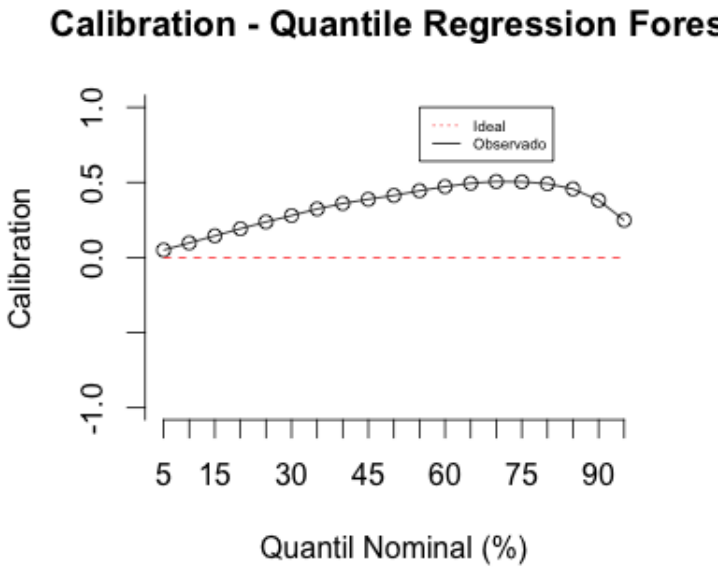

Figura 48: Quantile Regression Forests – Calibration

Como é possível observar na Figura 48, os valores de Calibration tem piores resultados entre os Quantis 60% e 85%. Porém ainda mostram um bom desempenho por não se afastarem muito do valor ideal.

Sharpness

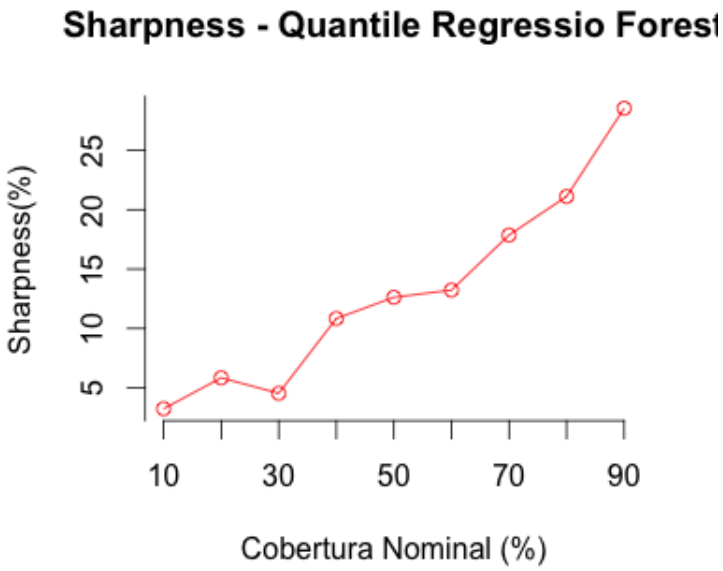

Figura 49: Quantile Regression Forests – Sharpness

Como mostra a Figura 49, a dispersão entre os quantis são maiores a medida que os valores dos quantis são aumentados. Porém há uma queda no valor de *Sharpness* no Quantil 30%.

Skill Score

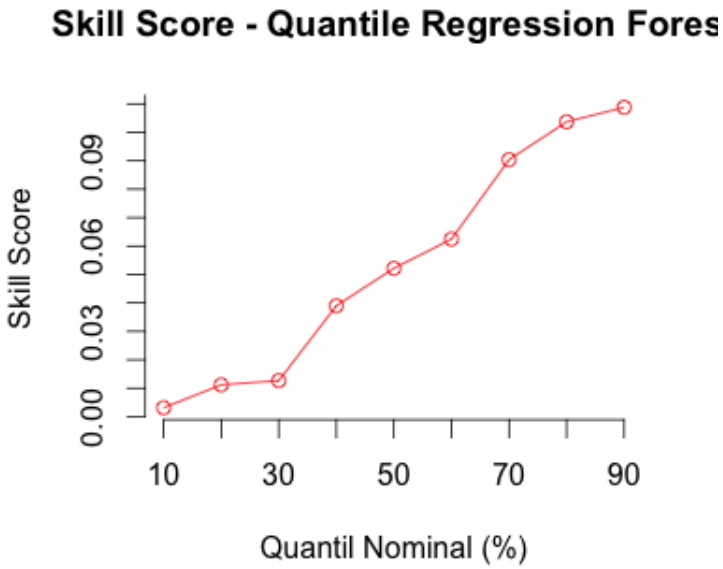

Figura 50: Quantile Regression Forests – Skill Score

É possível observar na Figura 50 que o *Skill Score* tem piores resultados a medida que os valores

dos Quantis são mais elevados, porém os valores são muito próximos de zero, mostrando um bom resultado.

Resumo dos Resultados de Desempenho para o Quantil 0.5:

MAE	RMSE	MAPE	Calibration	Sharpness	Skill Score
10.45725	13.47243	10.31639	0.4133418	0.1262564	0.05218704

Tabela 26: Desempenho Quantile Regression Forest (Quantil 0.5)

4.4.4 Resultados com Gradient Boosting Machines

4.4.4.1 Base de Dados Total

Com o package *gbm* [15] foi utilizado 5000 árvores conforme Figura 51.

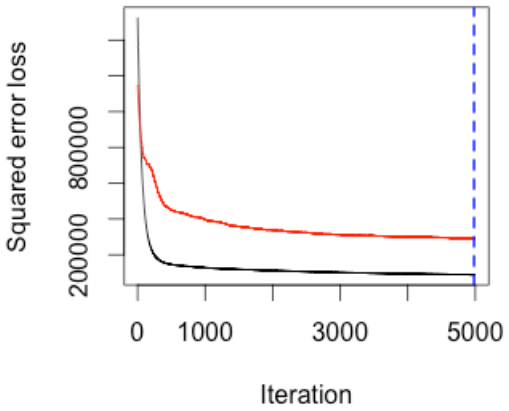

Figura 51: GBM – Número de árvores

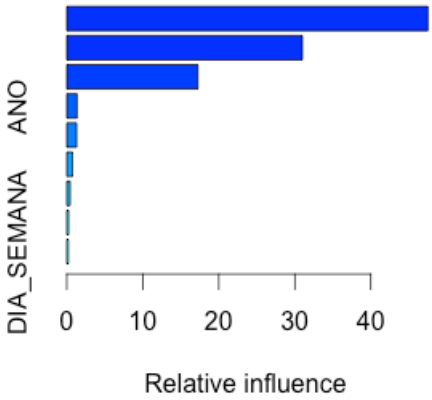

Figura 52: GBM – Influencia das Variáveis

Sumário:

Variável	Rel. Inf
PREV_CARGA	47.4957019
MES	30.9960044
PHF	17.2367000
ANO	1.3471622
DIA_MES	1.2920364
HORA	0.7623858
PREV_EOLICA	0.4386652
PREV_SOLAR	0.2329459
DIA_SEMANA	0.1983981

Tabela 27: Desempenho Quantile Regression Forest (Quantil 0.5)

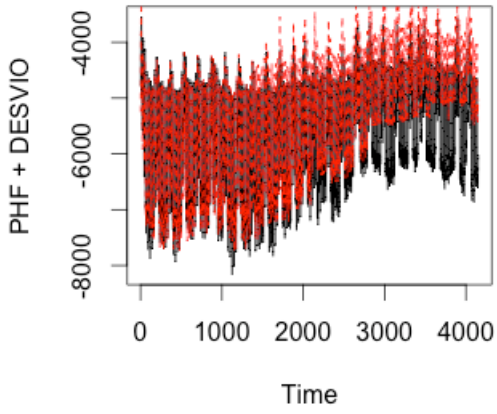

Figura 53: GBM – Valores Reais X Valores Previstos

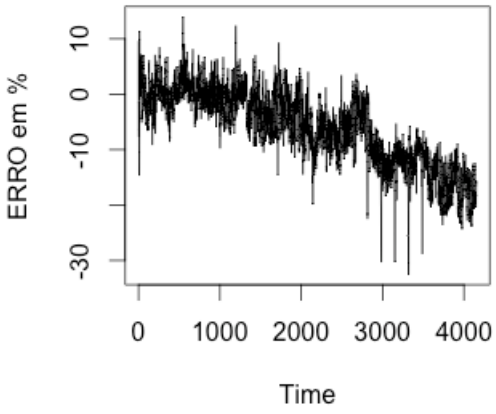

Figura 54: GBM – Erro

Calibration

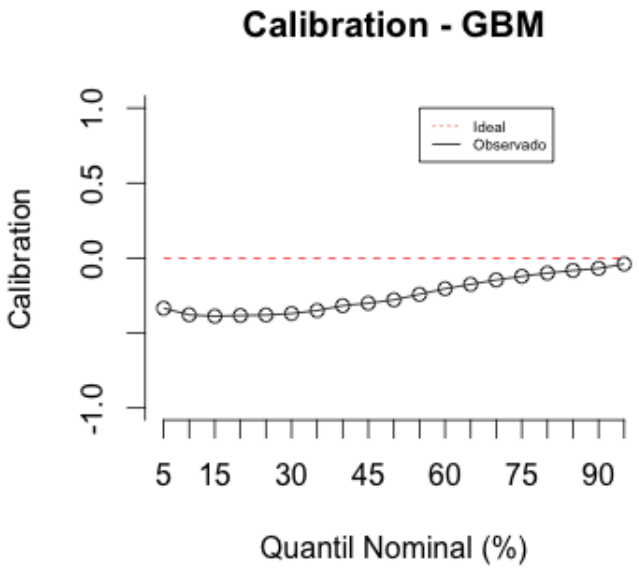

Figura 55: GBM – Calibration

Sharpness

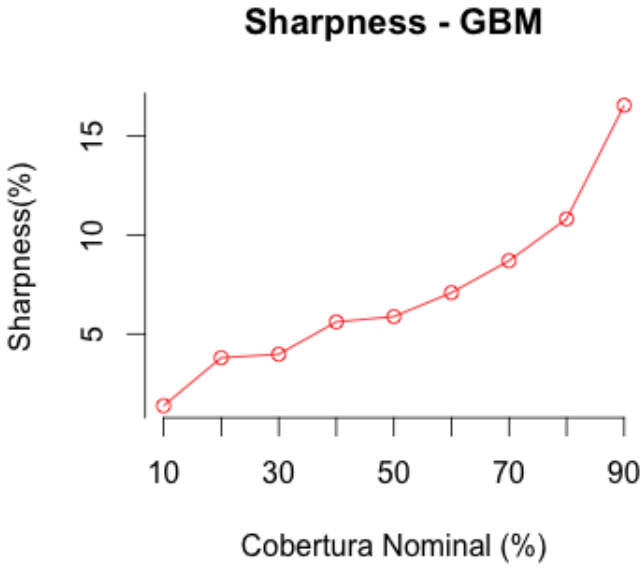

Figura 56: GBM – Sharpness

Skill Score

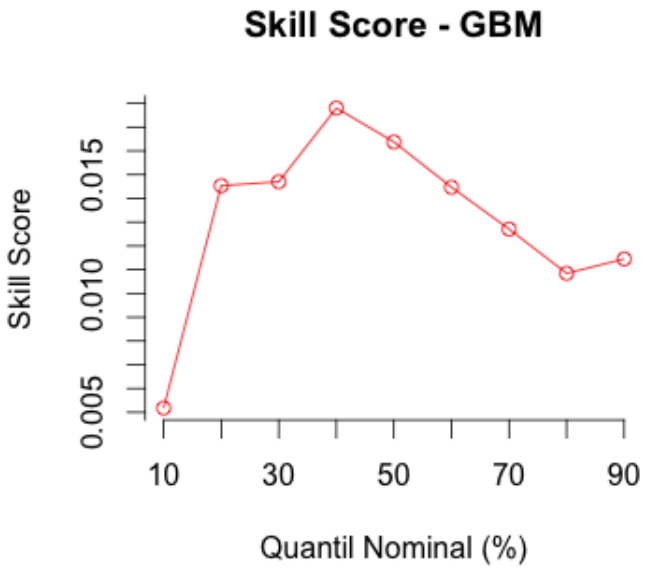

Figura 57: GBM – Skill Score

Resumo dos Resultados de Desempenho para o Quantil 0.5:

MAE	RMSE	MAPE	Calibration	Sharpness	Skill Score
5.584685	9.014268	5.775988	-0.2786609	0.05875563	0.0163729

Tabela 28: Desempenho Gradient Boosting Machine

4.4.4.2 Base de Dados AUDAC02

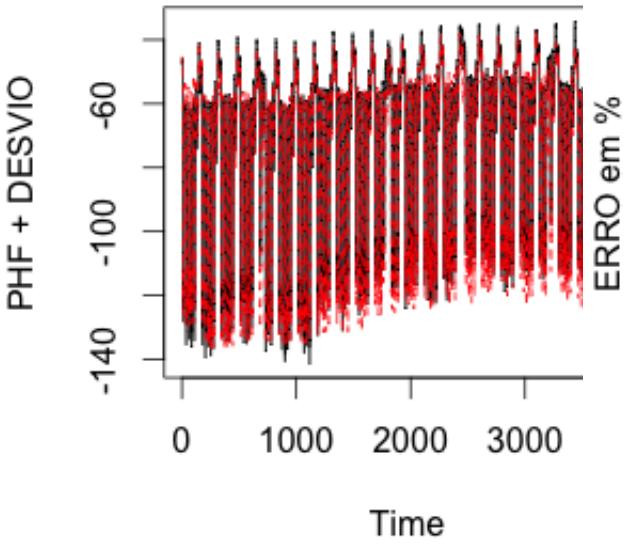

Figura 58: GBM – Número de árvores

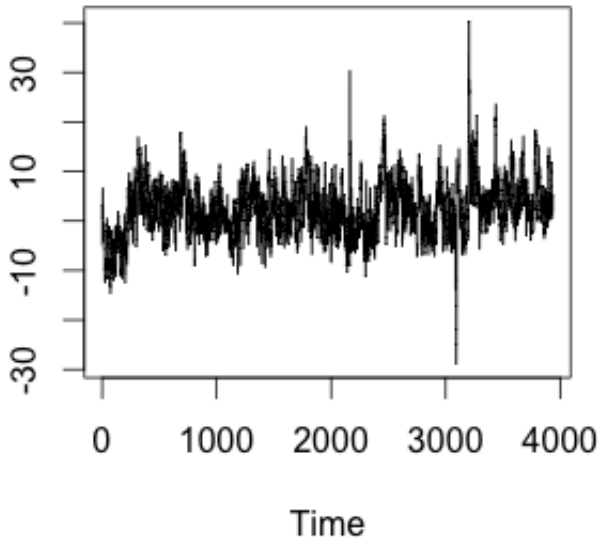

Figura 59: GBM – Influencia das Variáveis

Calibration

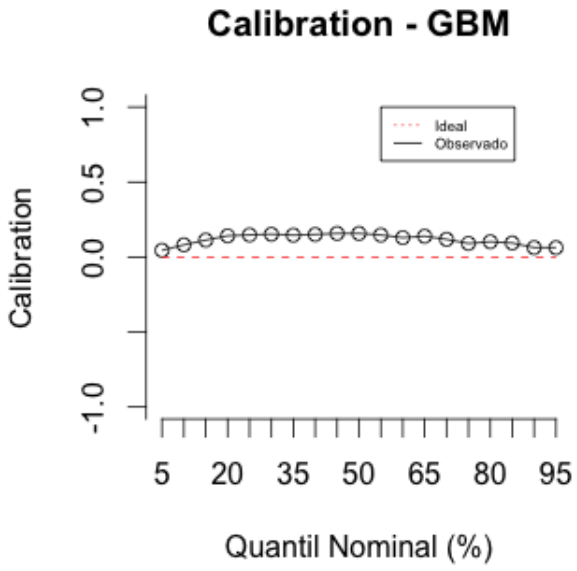

Figura 60: GBM – Calibration

Sharpness

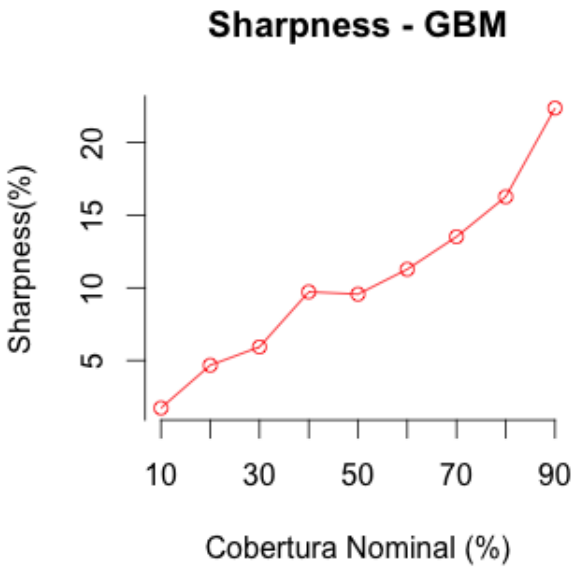

Figura 61: GBM – Sharpness

Skill Score

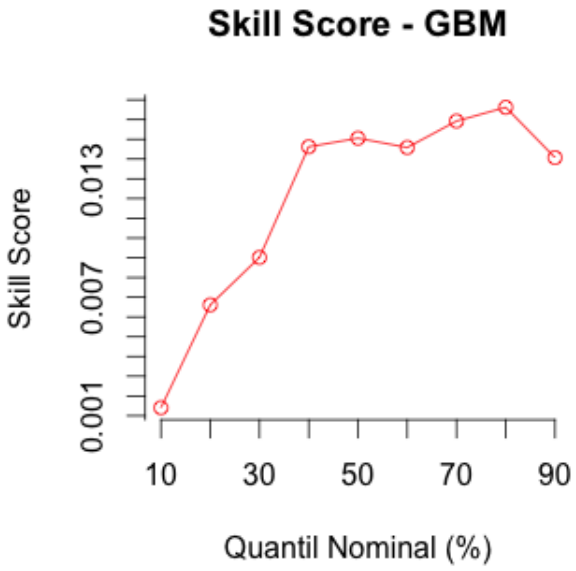

Figura 62: GBM – Skill Score

Resumo dos Resultados de Desempenho para o Quantil 0.5:

MAE	RMSE	MAPE	Calibration	Sharpness	Skill Score
2.210336	5.88234	2.330926	0.1574333	0.0956232	0.01505428

Tabela 29: Desempenho Gradient Boosting Machine

4.5 Comparação entre os Métodos - Base de Dados Total

Os três Métodos de Previsão Probabilística: Regressão Linear de Quantis, *Quantile Regression Forest* e *Gradient Boosting Machines* se mostraram eficientes para o problema apresentado.

Como não é tratado por Quantis, não é possível calcular *Calibration*, *Sharpness* e *Skill Score* para o método de Regressão Linear. Porém, para motivo de comparação dos Erros, foi acrescentado nas tabelas 30 e 31.

Calibration

Como é possível observar na Figura 63, o modelo com melhor desempenho em termos de *Calibration* é GBM que para qualquer Quantil Nominal possui valores de desvio mais próximos de zero do que os outros dois métodos.

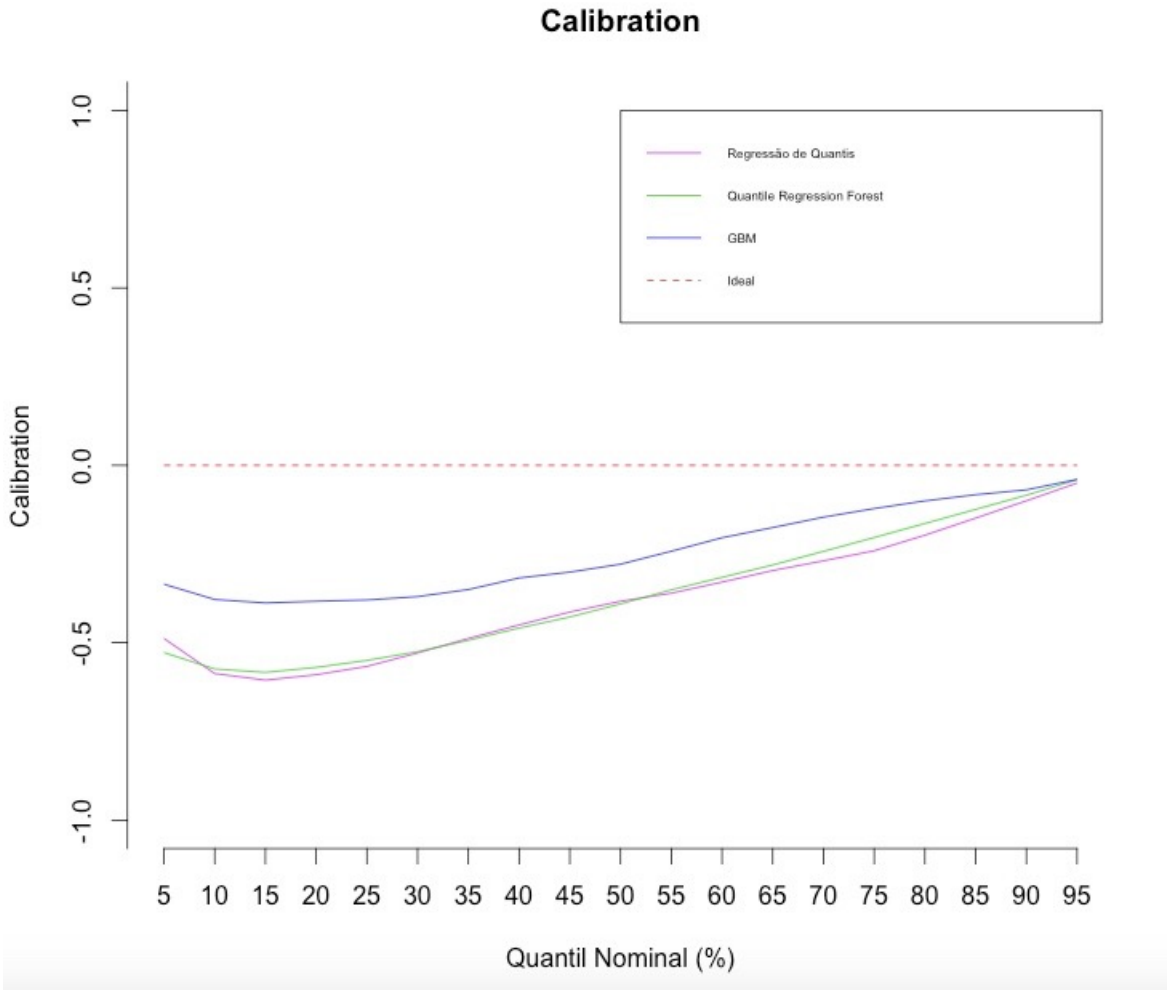

Figura 63: Comparação entre os Métodos: Calibration

Sharpness

Mais uma vez, o método GBM possui os melhores resultados. Os valores de Sharpness são

menores do que os outros métodos para todos os Quantis, significando que a dispersão entre os pares de Quantis são menores.

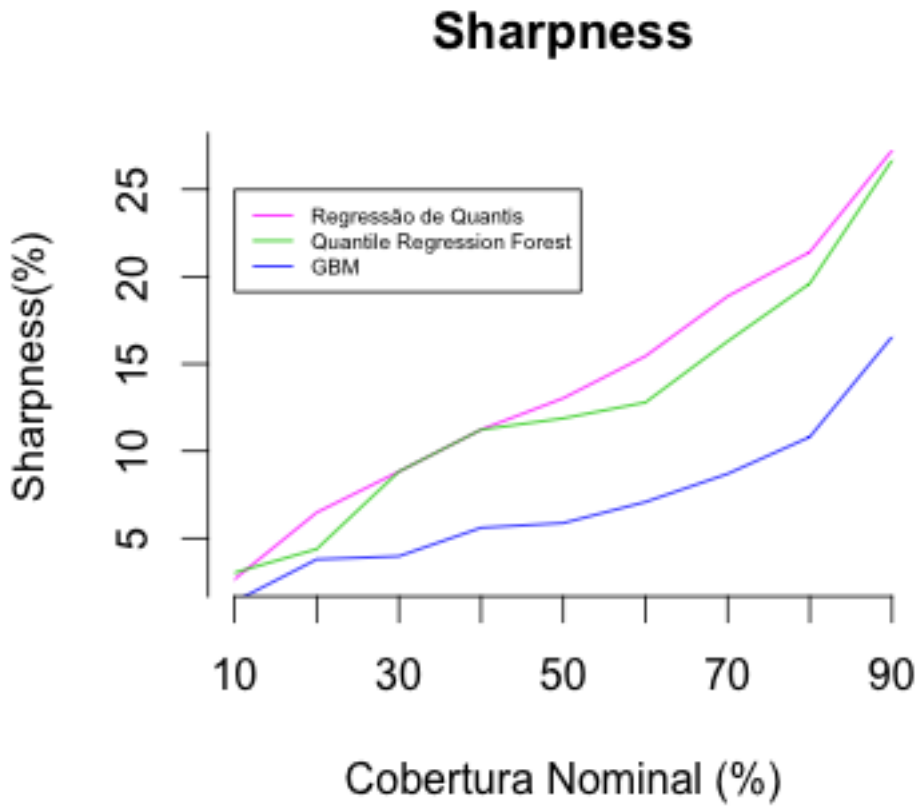

Figura 64: Comparação entre os Métodos: Sharpness

Skill Score

Consequentemente, o método de previsão probabilística com os melhores valores de *Skill Score* é o GBM, uma vez que o *Skill Score* possui informações tanto de *Calibration* quanto de *Sharpness*, e como visto anteriormente, o modelo de GBM apresentou melhores resultados nos dois métodos de avaliação.

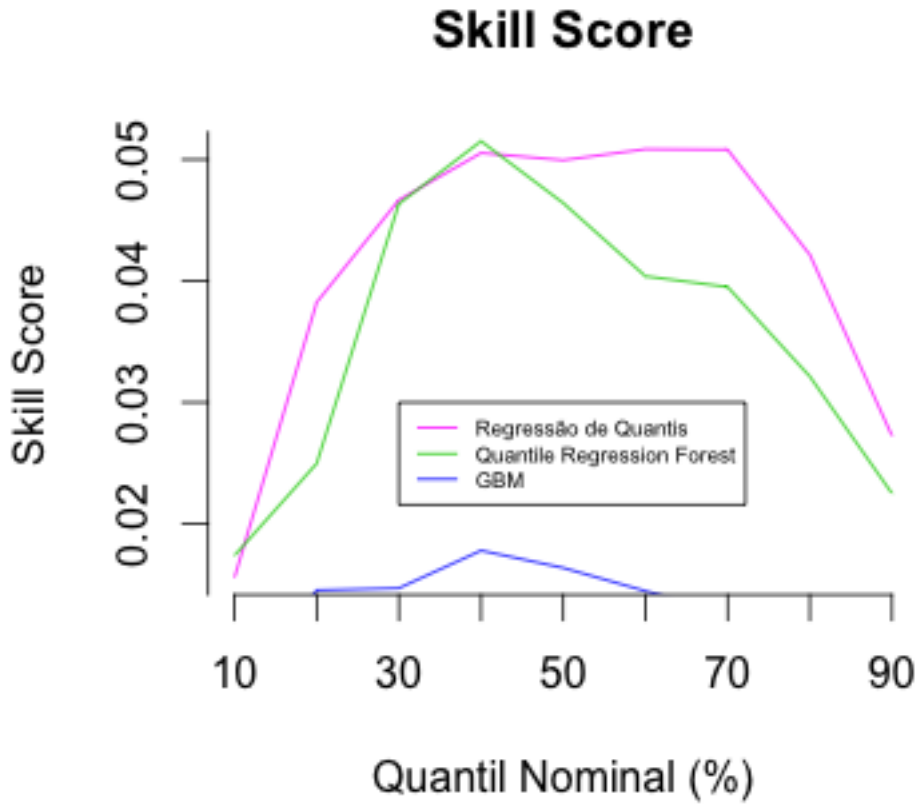

Figura 65: Comparação entre os Métodos: Skill Score

Modelo	MAE	RMSE	MAPE	Calibration	Sharpness	Skill Score
Regressão Linear	10.7474	11.71975	11.71975	-	-	-
Regressão de Quantis	8.0028	9.909769	8.121541	-0.3831888	0.130338	0.04994407
Quantile Regression Forest	10.70286	13.52778	11.29314	-0.3906551	0.1187999	0.04640977
GBM	5.584685	9.014268	5.775988	-0.2786609	0.05875563	0.0163729

Tabela 30: Comparação dos Resultados entre os Modelos

4.6 Comparação entre os Métodos - Base de Dados AUDAC02

Assim como na Base de Dados Total, os três Métodos de Previsão Probabilística: Regressão Linear de Quantis, *Quantile Regression Forest* e *Gradient Boosting Machines* se mostraram eficientes para o problema apresentado.

Calibration

Como é possível observar na Figura 66, o modelo com melhor desempenho em termos de *Cali-*

bration é GBM que para qualquer Quantil Nominal possui valores de desvio mais próximos de zero do que os outros dois métodos.

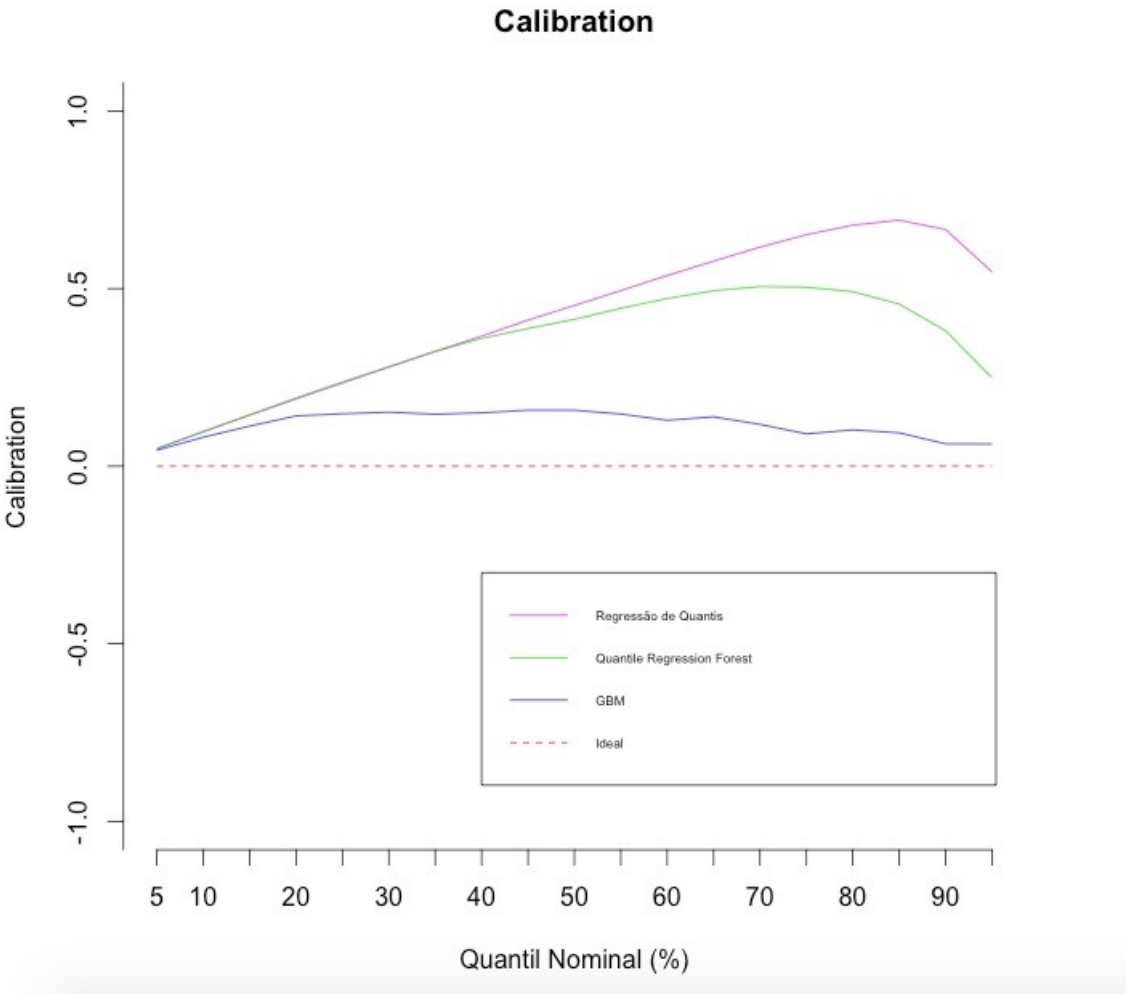

Figura 66: Comparação entre os Métodos: Calibration

Sharpness

Para a Base de Dados da unidade AUDAC02, o método com os melhores resultados de Sharpness foi Regressão Linear de Quantis.

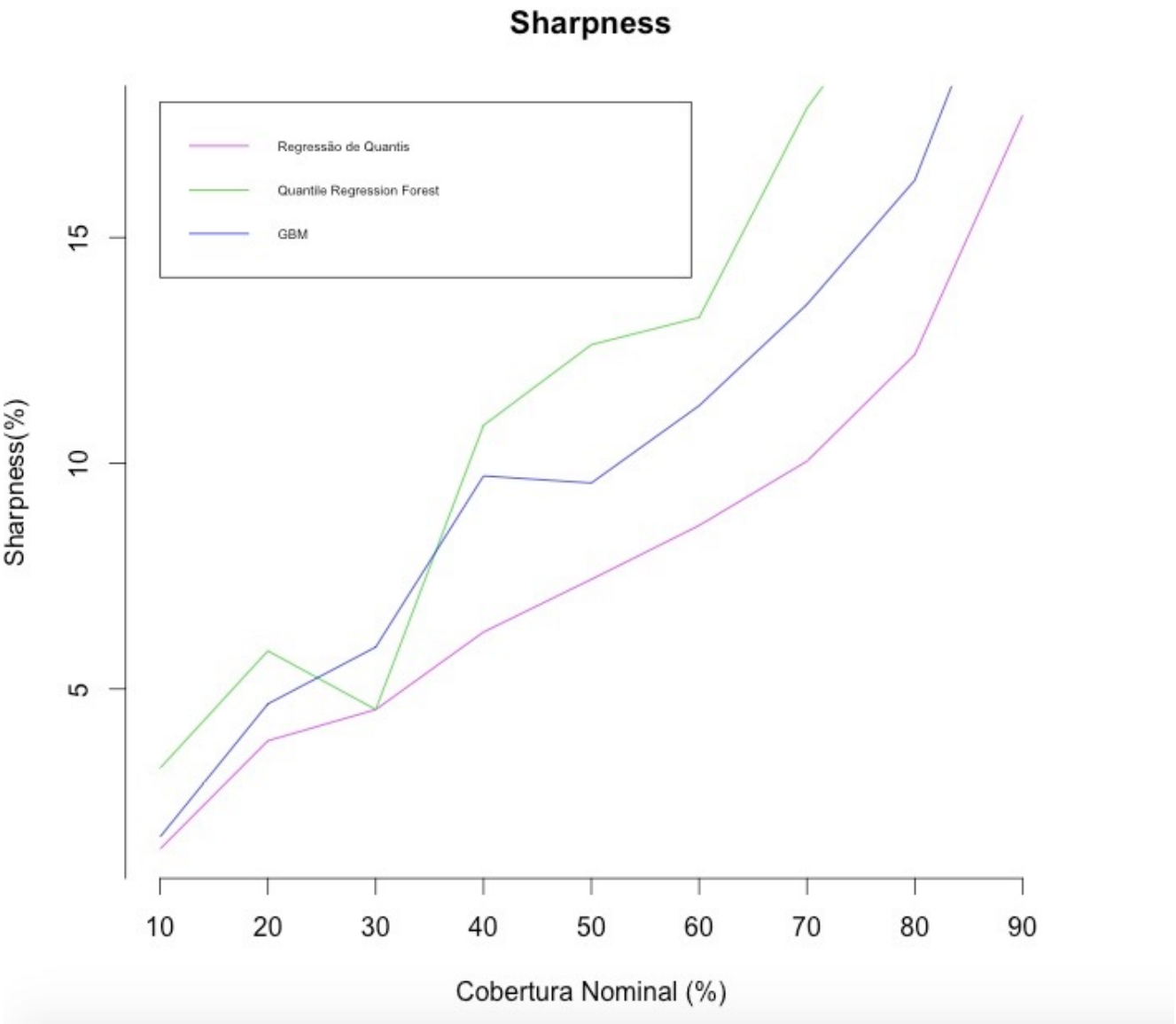

Figura 67: Comparação entre os Métodos: Sharpness

Skill Score

Assim como na Base de Dados Total, o método de previsão probabilística com os melhores valores de *Skill Score* é o GBM.

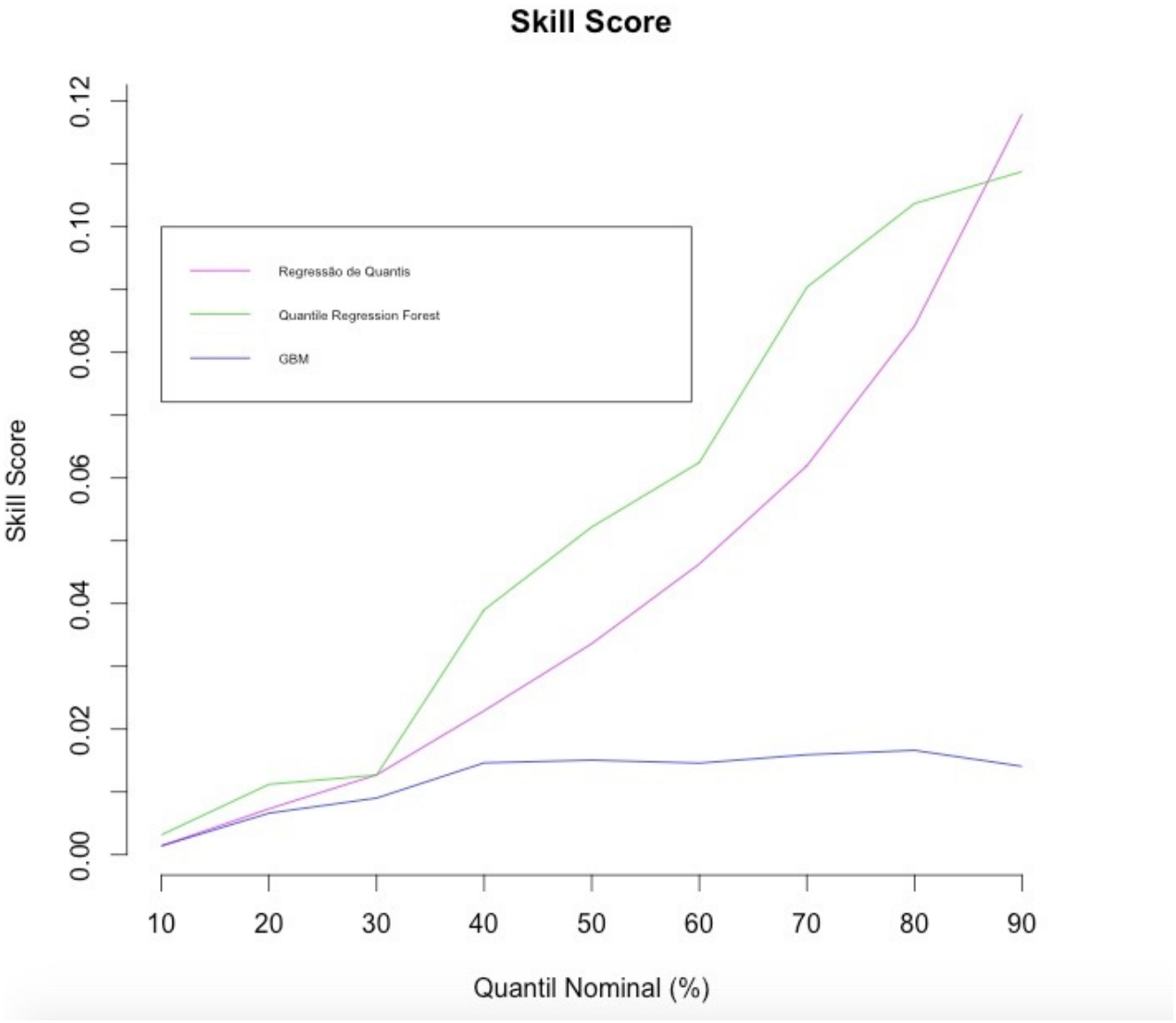

Figura 68: Comparação entre os Métodos: Skill Score

Modelo	MAE	RMSE	MAPE	Calibration	Sharpness	Skill Score
Regressão Linear	7.884847	9.375542	10.88228	-	-	-
Regressão de Quantis	10.16555	12.38017	10.23533	0.4524778	0.07428362	0.03361169
Quantile Regression Forest	10.45725	13.47243	10.31639	0.4133418	0.1262564	0.05218704
GBM	2.210336	5.88234	2.330926	0.1574333	0.0956232	0.01505428

Tabela 31: Comparação dos Resultados entre os Modelos

Conclusão

O cenário Europeu do Mercado de Energia Elétrica sofreu grandes mudanças ao longos dos últimos anos, e com essa reestruturação surgiu a necessidade de ampliar os estudos de previsão probabilística para esse mercado.

Com a criação do MIBEL, todos os agentes tiveram livre acesso a compra e venda de energia elétrica, ocasionando em uma maior volatilidade no mercado. A REN por sua vez necessita prever os desvios das unidades de cada agente, para isso é necessário a utilização de metodologias de previsão probabilística.

A presente dissertação explorou três métodos de previsão: Regresão Linear de Quantis, *Quantile Regression Forest* e *Gradient Boosting Machine*. Toda a análise de Erro, desvios e avaliações de desempenho que envolveram *Calibration*, *Sharpness* e *Skill Score* mostraram que para esse cenário o modelo de *Gradient Bosting Machine* tem a melhor performasse.

A Análise do Erro mostrou que o método de GBM possui a menor média de Erro associada. As análises de MAE, RMSE e MAPE para a Base de Dados Total, indicaram que os modelos de Regressão Linear, Regressão de Quantis e *Quantile Regression Forest* possuem valores próximos a 11% de Erro, enquanto o *Gradient Boosting Machine* possui valores próximos a 6%. Já para a Base de Dados da unidade AUDAC02 o valores de Erro são entre 2% e 6%.

As medidas de avaliação de desempenho *Calibration* e *Sharpness* para o Quantil 0.5 da Base de Dados Total indicaram o mesmo comportamento da análise do Erro. Enquanto a Regressão de Quantis e *Quantile Regression Forest* apresentaram valores próximos de -0.4 de *Calibration* e 0.12 de *Sharpness*, o metódo de GBM apresentou valores de aproximadamente -0.3 de *Calibration* e 0.06 de *Sharpness*. O que indica que os quantis estimados utilizando GBM diferem menos dos quantis nominais e que sua capacidade de prever acontecimentos de uma forma precisa é superior aos outros modelos apresentados.

Para a Base de Dados AUDAC02 os valores de *Calibration* foram melhores para o método de GBM, porém os valores de *Sharpness* foram melhores para o método de Regressão Linear de Quantis.

Como foi apresentado, com as medidas de *Calibration* e *Sharpness* separadamente não é possível avaliar a qualidade da previsão. Portanto, com a avaliação *Skill Score* é possível obter um resultado sobre a qualidade do método. E como essa avaliação consiste das informações de *Calibration* e *Sharpness*, o método de GBM obteve os melhores resultados tanto com a Base de Dados Total, quanto na Base de Dados AUDAC02. Enquanto o *Skill Score* dos métodos de Regressão de Quantis

e *Quantile Regression Forest* apresentaram valores próximos de 0.05, o método de GBM apresentou resultado próximo a 0.016. Um resultado consideravelmente melhor que os outros métodos.

Com a utilização dos métodos de previsão probabilística apresentados nessa dissertação para o problema apresentado dos desvios totais e dos agentes do mercado de eletricidade, e com as medidas de avaliação utilizadas, conclui-se que para esse cenário o método que melhor prevê o desvio é o *Gradient Boosting Machine*.

Trabalhos Futuros

Diferentes métodos de previsão probabilística se enquadram melhor em diferentes cenários, por esse motivo nesta dissertação foram aplicados três métodos. Porém, é necessário explorar mais métodos de previsão probabilística para o cenário de desvio de energia elétrica, tais como NW-KDE e Redes Neuronais Artificiais, e analisa-los se são bem enquadrados para esse problema.

A presença de *Missing Values* em variáveis significativas para os modelos podem ter um forte impacto na previsão. Nesta dissertação, foram excluídos os *Missing Values*, porém, a utilização de métodos de detecção, avaliação e simulação dos mesmos, poderá resultar em uma possível melhoria dos resultados.

Todas as variáveis explicativas apresentadas nesta dissertação se mostraram significativas para os modelos, porém outras variáveis podem também ter impacto nos desvios de energia elétrica, como por exemplo o clima, se o dia em questão é um feriado ou alguma data comemorativa, e mais fatores que possam influenciar a maior ou menor utilização de energia elétrica em determinado momento.

Referências

[1] Henrik Aalborg Nielsen*, Henrik Madsen and Torben Skov Nielsen, Using Quantil Regression to extend an existing wind power forecasting system with probabilistic forecasts, Technical University of Denmark, DK-2800 Lyngby, Denmark

[2] Moreira, Rui, Previsão Probabilística dos preços de energia elétrica do mercado Ibérico de eletricidade, Tese de Mestrado em Modelação, Analise de Dados e Sistemas de Apoio à Decisão, Faculdade de Economia da Universidade do Porto.

[3] Friedman, Jerome H., Greedy Function Approximation: A Gradient Boosting Machine, IMS 1999 Reitz Lecture, February 24,1999(modified March 15, 2000, April 19, 2001)

[4] Paula, Ebberth, Mineração de dados como suporte à detecção de lavagem de dinheiro, Tese de Mestrado em Computação Aplicada, Instituto de Ciências Exatas, Departamento de Ciência da Computação, universidade de Brasília.

[5] Koenker, R. and G. Bassett Jr, Regression quantiles. Econometrica: journal of the Econometric Society, 1978: p. 33-50.

[6] Oshiro, Thais, Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica, Tese de Mestrado em Bioinformática, Universidade de São Paulo.

[7] Meinshausen, N., Quantile regression forests. The Journal of Machine Learning Research, 2006. 7: p. 983-999.

[8] Bessa, Ricardo, Methodologies for the participation of an electric vehicle's aggregator in the electricity markets, Thesis submitted to the Faculty of Engineering of University of Porto

[9] Ribeiro, Luís, Previsão Probabilistica de preços de eletricidade para o mercado diário MIBEL, Tese de Mestrado em Engenharia Eletrotécnica e de Computadores, Faculdade de Engenharia da Universidade do Porto.

[10] Lopes, Silvia, Desenvolvimento de um modelo de previsão de preços no mercado Ibérico de eletricidade, Tese de Mestrado em Engenharia Eletrotécnica e de Computadores, Instituto Superior de Engenharia do Porto.

[11] REN – Sistema de Informação de Mercados de Energia. Disponível em: http://www.mercado.ren.pt/ . Acesso em 10 de Agosto de 2016.

[12] Entsoe – Tranparency Platform. Disponível em: https://transparency.entsoe.eu/ . Acesso em 10 de Agosto de 2016.

[13] Pierre Pinson*, Henrik Aa. Nielsen, Jan K. Møller and Henrik Madsen, Non-parametric Probabilistic Forecasts of Wind Power: Required Properties and Evaluation, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark

[14] Girard, Robin, Deliverable D-1.3 Towards the definition of a standardized evaluation protocol, v1.4 - 2009-01-07

[15] Ridgeway Greg, Generalized Boosted Regression Models, 2017-03-21

[16] Meinshausen Nicolai, Quantile Regression Forests, 2016-05-19

[17] Oliveira, Felipe António Sobral Sacramento, Previsão Probabilística dos Preços de Eletricidade, Dissertação realizada no âmbito do Mestrado Integrado em Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia da Universidade do Porto, 2015.

[18] Friedman, J. H. Stochastic Gradient Boosting, Universidade de Stanford, 1999.

[19] Guimarães, Ana Catarina Fontes, Previsão da evolução da carga no médio prazo em redes de distribuição, Dissertação para obtenção do Grau de Mestre em Engenharia Eletrotécnica e de Computadores, Instituto Superior Técnico, Universidade Técnica de Lisboa, 2008.

[20] R. Guimarães, J. Cabral, "Estatística", Mc Graw Hill, 1997.

[21] P. Pinson, G. Kariniotakis, H. Aa. Nielsen†, T. S. Nielsen, H. Madsen, Properties of Quantile and Interval Forecasts of Wind Generation and their Evaluation

[22] Mark P.J. van der Loo, Edwin de Jonge, Learning RStudio for R Statistical Computing, December 2012

[23] Gneiting, Tilmann, Balabdaoui, Fadoua and Raftery, Adrian E., Probabilistic forecasts, calibration and sharpness, J. R. Statist. Soc. B (2007)

[24] Gomes, A., Previsão a Curto Prazo dos Preços de Mercado Diário de Eletricidade, Dissertação de Mestrado Integrado em Engenharia Electrotécnica e Computadores, FEUP, Julho de 2014

[25] Ribeiro, J., Previsão de preços de eletricidade para o mercado MIBEL, Dissertação de Mestrado Integrado em Engenharia Electrotécnica e Computadores, FEUP, Junho de 2014

[26] Tomé, B., Previsão de Preços de Energia Eléctrica em Mercados de Eletricidade – Horizonte de 24 Horas, Dissertação de Mestrado Integrado em Engenharia Electrotécnica e Computadores, FEUP, Junho de 2009

[27] Duarte, A., Previsão de Preços de Energia Eléctrica em Mercados de Electricidade – Horizonte de uma semana, Dissertação de Mestrado Integrado em Engenharia Electrotécnica e Computadores, FEUP, Junho de 2008

[28] Gomes, G., Previsão a longo prazo de preços de electricidade, Dissertação de Mestrado Integrado em Engenharia Electrotécnica e Computadores, FEUP, Fevereiro de 2010

[29] Ferreira, José, Modelos de Regressão para Previsão de Sinistros, Dissertação de Mestrado em Engenharia Matemática, FCUP, Setembro de 2013

[30] Miranda, Carla, Modelação Linear de Séries Temporais na presença de Outliers, Dessertação de Mestrado em Estatística, FCUP, Março de 2001.